MAT 1100 - ALGEBRA I

CHRISTOPHER EAGLE

Abstract. These are my notes from the MAT 1100 lectures given by Prof. Bar-Natan in Fall 2010 at the University of Toronto.

1. Groups

1.1. **Introduction - The Rubik's Cube.**

Notes from September 14 should go here, but are posted by someone else on the website.

1.2. **The Basics.**

We have seen what groups are, and wish to study them in more detail. No group exists all alone, however, and it is useful to define structure-preserving maps between groups.

Definition 1.1. Let G, H be groups. A **homomorphism** (or just **morphism**) is a function $\phi : G \to H$ which preserves the group structure of G, in the sense that:

(1) For all $g_1, g_2 \in G$, $\phi(g_1 g_2) = \phi(g_1)\phi(g_2)$.
(2) $\phi(e_G) = e_H$.
(3) For all $g \in G$, $\phi(g^{-1}) = \phi(g)^{-1}$.

A **isomorphism** is a bijective homomorphism. If there is some $\phi : G \to H$ which is an isomorphism, we write $G \cong H$.

We note that points (2) and (3) in the above definition actually follow from point (1), and hence need not be checked independently when verifying that a map is a homomorphism. Indeed, using (1) we have $e_H \phi(e_G) = \phi(e_G e_G) = \phi(e_G)\phi(e_G)$, so by the cancelation law $\phi(e_G) = e_H$. Also, for any $g \in G$, we have $e_H = \phi(e_G) = \phi(gg^{-1}) = \phi(g)\phi(g^{-1})$, and hence $\phi(g^{-1}) = \phi(g)^{-1}$ by definition of inverses. We also note that $\phi(g^n) = \phi(g)^n$.

The universe (Groups, morphisms) forms an example of a "Category". We will encounter categories in more detail later. In particular, the definition of a category will include the following:

(1) Morphisms can be composed, and the result is again a morphism.
(2) Every object (i.e., every group) has a distinguished "identity morphism".

That is, if we have three groups G, H, K, and morphisms $\phi : G \to H$ and $\psi : H \to K$, then the composition $\psi \circ \phi : G \to K$ is a morphism of groups, and there is an identity map $I_K : K \to K$, namely for all $k \in K$, $I_K(k) = k$. The key property of I_K is that it is a morphism, and for any morphism $\phi : G \to K$ we have $I_K \circ \phi = \phi$. All of the above are trivial to check.

Example 1.2. (1) If V, W are vector spaces then they are, in particular, groups under addition. A linear transformation $T : V \to W$ is in particular a group morphism.

Date: December 8, 2010.

(2) The map $\exp : (\mathbb{R}, +) \to (\mathbb{R}^+, \times)$ given by $x \mapsto e^x$ is a group morphism.
(3) If $H \leq G$ then the inclusion map $i_H : H \to G$, given by $i_H(h) = h$ for every $h \in H$, is a group morphism.
(4) Given any group G, and any $g \in G$, define conjugation by g to be the map (from G to G) given by $h \mapsto h^g = g^{-1}hg$. Then:
  (a) $(h_1h_2)^g = h_1^g h_2^g$, so conjugation is a morphism. As it maps G to itself, it is an **endomorphism**.
  (b) $h^{g_1g_2} = (h^{g_1})^{g_2}$. It follows that the conjugation map is invertible, and $(h^g)^{g^{-1}} = h$. So conjugation is an **automorphism**.
  (c) $(a^b)^c = (a^c)^{(b^c)}$. Why this is important will become clear later.
  (d) If ϕ is a homomorphism then ϕ preserves conjugation, in the sense that $\phi(h^g) = \phi(h)^{\phi(g)}$.
(5) Consider S_4 as the automorphism group of the tetrahedron (this autmorphism group really is S_4, since any vertex of the tetrahedron can be mapped to any other, and then any remaining vertex mapped to any remaining vertex, and so forth). Consider S_3 as the symmetric group on the colours R, G, B. We define a (surjective) morphism $\phi : S_4 \to S_3$ by mapping pairs of opposite (i.e., non-intersecting) edges the same colour. Since an automorphism of a tetrahedron preserves the adjacency relation among edges, but might change the colours of the pairs. Thus ϕ is indeed a morphism.

    For example, consider the permutation $\sigma = 2341 \in S_4$. Then $\phi(2341)$ is $R \mapsto G$, $G \mapsto R$, and $B \mapsto B$. [insert picture]

Claim 1.3. If $\phi : G \to H$ is a morphism then $\ker \phi := \phi^{-1}(e_H) \leq G$ and $\operatorname{im} \phi = \phi(G) \leq H$.

We have seen in the above example that S_3 is an image of S_4. Note that $S_3 \leq S_4$ is clear. Is S_3 also the kernel of some $\phi : S_4 \to G$ for some G? The answer is no.

Suppose that $\phi : G \to H$ is a homomorphism. Note that if $h \in \ker \phi$ then for any $g \in G$ we have $\phi(h^g) = \phi(h)^{\phi(g)} = e_H^{\phi(g)} = e_H$, so $h^g \in \ker \phi$. This is an example of the following definition:

Definition 1.4. A subgroup $N \leq G$ is called **normal** if for every $g \in G$ and every $n \in N$, $n^g \in N$. We denote this by $N \trianglelefteq G$.

Above, we thus proved the following:

Proposition 1.5. *If $\phi : G \to H$ is a homomorphism, then $\ker \phi \lhd G$.*

So to answer our question about S_3, we should check if $S_3 \lhd S_4$. Consider the permutation $2314 \in S_3 \leq S_4$. If we conjugate by something in S_3 we know we will be back in S_3, so we try conjugation by $1243 \in S_4 \setminus S_3$. We get $[1243]^{-1}[2314][1243]$ sends 4 to 1, so the conjugation is not in S_3. Thus $S_3 \ntrianglelefteq S_4$, so S_3 cannot be the kernel of any morphism from S_4.

As an exercise, compute the kernel of each homomorphism in Example 1.2. In particular, the kernel in part (5) is non-trivial, having 4 elements.

We now ask about the converse of Proposition 1.5. That is, given $N \trianglelefteq G$, is there a surjective morphism $\phi : G \to H$ such that $N = \ker \phi$? The answer is yes, and we will prove it after a set-theoretic aside.

Recall that an **equivalence relation** on a set X is a binary relation $\sim$ on X such that:

1. **Reflexive:** $x \sim x$
2. **Symmetric:** $x \sim y \iff y \sim x$
3. **Transitive:** $x \sim y, y \sim z \implies x \sim z$

Given an equivalence relation $\sim$ on X we get a new set $X/\sim = \{[x]_\sim : x \in X\}$ where we define $[x]_\sim = \{y \in X : y \sim x\}$. X is thus decomposed into a disjoint union of its equivalence classes under $\sim$. We recall that equivalence relations are closely related to surjections. Given an equivalence relation $\sim$, we get a surjective map $\pi : X \to X/\sim$ given by $\pi(x) = [x]_\sim$. Conversely, given a surjective map $\phi : X \to Y$, we get an equivalence relation on X by defining $x_1 \sim x_2 \iff \phi(x_1) = \phi(x_2)$. These operations define a "natural equivalence of categories", which we will not define.

We can now return to our group-theoretic problem. If ϕ existed, and $\sim$ was the corresponding equivalence relation, what properties would $\sim$ have? Note that:

$$\begin{aligned} g_1 \sim g_2 &\iff \phi(g_1) = \phi(g_2) \\ &\iff \phi(g_1^{-1} g_2) = e_H \\ &\iff g_1^{-1} g_2 \in N \qquad\qquad \iff g_2 \in g_1 N \end{aligned}$$

We can now use this thought process to resolve our question:

Proposition 1.6. *Let G be a group, $N \trianglelefteq G$ a normal subgroup. Then there is a group H and a morphism $\phi : G \to H$ such that $N = \ker \phi$.*

Proof. Define $\sim$ on X by $g_1 \sim g_2 \iff g_1^{-1} g_2 \in N (\iff g_2 \in g_1 N)$. We claim that $\sim$ is an equivalence relation. The proofs that $\sim$ is reflexive and symmetric are easy, and omitted. Suppose that $g_1 \sim g_2 \sim g_3$. Then $g_2 = g_1 n_1$ for some $n_1 \in N$ and $g_3 = g_2 n_2$ for some $n_2 \in N$. Then $g_3 = g_2 n_2 = g_1 n_1 n_2 \in g_1 N$ since $N \leq G$. So $g_1 \sim g_3$, and $\sim$ is an equivalence relation.

We now have the set $G/\sim$ and a surjective map $\phi : G \to G/\sim$ given by $g \mapsto [g]_\sim$. We write G/N instead of $G/\sim$, and $[g]_N$ (or just $[g]$) instead of $[g]_\sim$. If we can show that G/N is a group and ϕ is a morphism then we'll be done.

We have $[g_1][g_2] = \phi(g_1)\phi(g_2)$ by definition. As we want ϕ to be a morphism, we must define $[g_1][g_2] = [g_1 g_2]$, but we need to see that this is well-defined. Suppose $g_1 \sim g_1'$ and $g_2 \sim g_2'$. Then we must show that $g_1' g_2' \sim g_1 g_2$. Write $g_1' = g_1 n_1, g_2' = g_2 n_2$. Then $g_1' g_2' = g_1 n_1 g_2 n_2 = g_1 g_2 g_2^{-1} n_1 g_2 n_2 = g_1 g_2 n_1^{g_2} n_2 \in g_1 g_2 N$ since $N \trianglelefteq G$, so we got $g_1' g_2' \sim g_1 g_2$.

We next need to check that this multiplication makes G/N into a group, but this is trivial. Similarly, we also need to check that ϕ is a homomorphism, but we constructed it to be such, so this proof is also trivial. Also, that $N = \ker \phi$ is immediate from the definition of ϕ. □

We will frequently use the notion of G/N defined in the above proof. Along the way to the previous result, we also essentially proved the following theorem:

Theorem 1.7 (First Isomorphism Theorem)**.** *Let G, H be groups, $\phi : G \to H$ a homomorphism. Then $G/\ker \phi \cong \operatorname{im} \phi$.*

Proof. Define $\psi : G/\ker \phi \to \operatorname{im} \phi$ by $\psi([g]_{\ker \phi}) = \phi(g)$. It is a straightforward exercise to verify that:

1. ψ is well-defined.
2. ψ is a morphism.
3. ψ is bijective.

□

Even when $H \leq G$ but $H \ntrianglelefteq G$ we still have the following facts:

Lemma 1.8. *There is a bijection $[g_1] \to [g_2]$ for any $g_1, g_2 \in G$.*

Proof. We leave it as an exercise to check that $g_1' \mapsto g_2 g_1^{-1} g_1'$ for any $g_1' \in [g_1]$ is such a bijection. □

Corollary 1.9. *All equivalence classes mod H have the same size, namely $|H|$.*

Proof. This is immediate from the above and the fact that $H = [1_G]$. □

The above results show that $|G| = k\,|H|$ for some k. If G is finite, then $|H| \mid |G|$. We write $[G : H] = |G/H| = |G| \,/\, |H|$, and call this number the **index** of H in G. Note that sometimes we write $\overline{g}$ or gH for $[g] \in G/H$ when $H \leq G$.

Read along: Selick's notes 1.1, 1.2.1, 1.4. Lang's book $I1 - 3$.

Our goal is to reach the Jordan-Hölder Theorem, which asserts that every finite group G can be written (essentially uniquely) as a tower of normal extensions $G = G_0 \triangleright G_1 \triangleright G_2 \triangleright \ldots \triangleright G_n = \{1\}$, and this can be done in such a way that the sequence cannot be refined further. That is, for each i, G_i/G_{i+1} is **simple** - it has no non-trivial normal subgroups. We think of simplicity as a kind of primeness condition for groups, so the Jordan-Hölder theorem is a kind of prime decomposition for finite groups. Before we can get to this result, we will need several more isomorphism theorems.

Definition 1.10. Suppose that $K \leq G$. The **normalizer of** K **in** G is $N_G(K) = \left\{g \in G : g^{-1}Kg = K\right\}$.

Proposition 1.11. *Let $K \leq G$ be groups. Then*

(1) $N_G(K)$ is a group.
(2) $K \trianglelefteq N_G(K)$.
(3) $N_G(K) = G \iff K \trianglelefteq G$.

Proof. All are easy exercises. □

Theorem 1.12 (Second Isomorphism Theorem)**.** *Suppose that $H, K \leq G$ and $H \leq N_G(K)$. Then $H \cap K \trianglelefteq H$, $K \trianglelefteq H \cdot K$ (in particular, $H \cdot K$ is a group), and $H/(H \cap K) \cong (H \cdot K)/K$.*

(see picture)

Proof. We list the things that need checking, and observe that they are all easy.

- $H \cap K$ is a group.
- $H \cap K \trianglelefteq H$.
  - Take $h \in H$. Then for any $k \in H \cap K$, $h^{-1}kh \in H$ since H is a group, and $h^{-1}kh \in K$ since $H \leq N_G(K)$. So $h^{-1}kh \in H \cap K$.
- $H \cdot K$ is a group.
  - Take $h_1k_1, h_2k_2 \in H{\cdot}K$. Then $h_1k_1h_2k_2 = h_1h_2h_2^{-1}k_1h_2k_2 = h_1h_2(k_1)^{h_2}k_2 \in H \cdot K$ since $H \leq N_G(K)$. Inverses are similar.
- $K \trianglelefteq H \cdot K$.
  - Take $k_1 \in K$, $hk_2 \in H \cdot K$. Then $k_1^{hk_2} = (k_1^h)^{k_2} \in K$ since $k_1^h \in K$, as $H \leq N_G(K)$.

Define $\phi : H/(H \cap K) \to (H \cdot K)/K$ by $\phi([h]_{H \cap K}) = [h]_K$.

- ϕ is well-defined.
  - Consider ht with $t \in H \cap K$. Then $\phi([ht]) = [ht]_K = [h]_K = \phi([h]_{H\cap K})$ since $t \in K$.
- ϕ is a group morphism.

Define $\psi : (H \cdot K)/K \to H/(H \cap K)$ by $\psi([hk]_K) = [h]_{H\cap K}$.

- ψ is well-defined.
- ψ is a group morphism.
- ψ and ϕ are inverses of each other.

□

Theorem 1.13 (Third Isomorphism Theorem)**.** *Suppose $H \trianglelefteq G$, $N \trianglelefteq G$, and $N \leq H$ (so $N \trianglelefteq H$ as well). Then $H/N \trianglelefteq G/N$, and $(G/N)/(H/N) \cong G/H$.*

Proof. It is easy to check that $H/N \trianglelefteq G/N$. Define $\phi : (G/N)/(H/N) \to G/H$ by $\phi([[g]_N]_{H/N}) = [g]_H$. We first show that ϕ is well-defined. Observe that $[[gn]_N]_{H/N} \mapsto [gn]_H = [g]_H$ since $N \leq H$. We also must check well-definedness at the level of H/N, but this is similarly not difficult. Also easy to see is that ϕ is a group morphism.

Define $\psi : G/H \to (G/N)/(H/N)$ by $\psi([g]_H) = [[g]_N]_{H/N}$. Again, one easily checks that this is well-defined, a group morphism, and the inverse to ϕ. □

The next theorem says that given $N \trianglelefteq G$, then the subgroup lattice of G over N is the same as the subgroup lattice of G/N over $\{1\}$.

Theorem 1.14 (Fourth Isomorphism Theorem)**.** *If $N \trianglelefteq G$ then there is a bijection between subgroups of G that contain N and subgroups of G/N. This bijection preserves the notions of subgroup, indices, and intersections.*

Proof. Given $N \leq H \leq G$, since $N \trianglelefteq G$ we have $N \trianglelefteq H$. The bijection is $H \mapsto N/H$. The details are omitted. □

Lemma 1.15 (Butterfly Lemma)**.** *It $\{e\} < a \triangleleft A < G$ and $\{e\} < b \triangleleft B < G$, then $a(A \cap B)/a(A \cap b) \cong (A \cap B)b/(a \cap B)b$. See also the picture. Informally, "B/b", viewed in the "$a - A$ scale", is isomorphic to "A/a" viewed in the "$b - B$ scale".*

Our goal will be to prove the Jordan-Hölder Theorem, which is a sort of "prime decomposition" for groups. We will do this first, then return to prove the Butterfly Lemma. We will first need some definitions:

Definition 1.16. A group G is called **simple** if it has no non-trivial normal subgroups. That is, the only $N \trianglelefteq G$ are $N = \{e\}$ and $N = G$.

Definition 1.17. A **normal tower** for a group G is a sequence $G = A_n \triangleright A_{n-1} \triangleright \ldots \triangleright A_1 \triangleright A_0 = \{e\}$. Such a tower is called a **composition series** if A_k/A_{k-1} is simple for all k. We often write (A) for the sequence above.

Proposition 1.18. *Let G be a finite group. Then G has a composition series.*

Proof. Start with the tower $G \triangleright \{e\}$. If G is simple we're done, otherwise find a normal subgroup in between, insert it, and start again. Continue iterating. Since G is finite, the process must terminate, and hence produces a composition series. □

Definition 1.19. Two normal towers for G, say $(B)_0^m$ and $(A)_0^n$ are called **equivalent** if $n = m$ and there is $\sigma \in S_n$ such that $A_k/A_{k-1} \cong B_{\sigma(k)}/B_{\sigma(k)-1}$.

In general, a group will have many inequivalent normal towers. For composition series, however, this does not happen, as we shall see shortly.

Definition 1.20. A tower (C) is a **refinement** of a tower (A) if (A) comes from (C) by dropping some of the entries in the tower.

Theorem 1.21 (Jordan-Hölder)**.** *If $G = A_n \rhd A_{n-1} \rhd \ldots \rhd A_1 \rhd A_0 = \{e\}$ is a composition series for G, and $G = B_m \rhd B_{m-1} \rhd \ldots \rhd B_1 \rhd B_0 = \{e\}$ is another composition series for G, then the two series are equivalent.*

Once we have this theorem the study of groups (at least, finite groups) will be reduced to the study of simple groups and normal extensions. Before we can prove Jordan-Hölder, we will need a preliminary result.

Proposition 1.22. *Any two normal towers (A) and (B) for G have equivalent refinements.*

Proof of Theorem 1.21 from Proposition 1.22. Any refinement of a composition series is the series itself, by definition of composition series. By Proposition 1.22 the composition series (A) and (B) have equivalent refinements, which must be themselves, so (A) and (B) must be already equivalent. □

Proof of Proposition 1.22 from Lemma 1.15. Suppose $G = A_3 \rhd A_2 \rhd A_1 \rhd A_0 = \{e\}$, and $G = B_2 \rhd B_1 \rhd \{e\}$. We will use Lemma 1.15 to put $A_{1,1}$, which will be the image of B_1 under the compression, between A_0 and A_1. Then we will do this again to get $A_{2,1}$ between A_1 and A_2, and again to put $A_{3,1}$ between A_2 and A_3. This gives a refinement of (A), call it (A'). Now we do the reverse, putting $B_{1,1}$ and $B_{1,2}$ (images of A_1 and A_2, respectively, under compression) between B_0 and B_1. Similarly, put $B_{2,1}$ and $B_{2,2}$ between B_1 and B_2. This gives a refinement (B') of (B). Both $q(A')$ and (B') have $6 = 2 \cdot 3 = mn$ entries. The Butterfly Lemma 1.15 then says that the appropriate comparisons are isomorphic (see picture).

Now we give the formal proof. Let $A_{k,l} = A_{k-1}(B_l \cap A_k)$. Likewise, let $B_{l,k} = (B_l \cap A_k)B_{l-1}$. Then the sequence $(A_{k,l})$ (ordered $A_{n,m}, A_{n,m-1}, \ldots, A_{n,0}, A_{n-1,m}, \ldots$) refines (A_k) and $(B_{l,k})$ (ordered likewise) refines (B_l). By the Butterfly Lemma 1.15 $A_{k,l}/A_{k,l-1} \cong B_{l,k}/B_{l,k-1}$, where we understand $A_{k,l-1}$ to be $A_{k-1,m}$ if $l = 0$, and $B_{l,k-1} = B_{l-1,n}$ if $k = 0$. This causes no problems, since we always have $A_{k,0} = A_{k-1,m} = A_{k-1}$ and $B_{l,0} = B_{l-1,n} = B_{l-1}$. □

To complete our proof of the Jordan-Hölder Theorem, we now need only to prove the Butterfly Lemma

Proof of Lemma 1.15. Recall we have $\{e\} < a \lhd A < G$ and $\{e\} < b \lhd B < G$.

(1) Why is $a(A \cap B)$ a group? This, and all of the other statements of this form, follows from the Second Isomorphism Theorem 1.12 and the fact that $A < N_G(a)$, and so $A \cap B < N_G(a)$.
(2) $a(A \cap B) \rhd a(A \cap b)$. Take $\alpha_1 \in a, \alpha_2 \in A \cap B$, so $\alpha_1\alpha_2 \in a(A \cap B)$. Then $(a(A \cap b))^{\alpha_1\alpha_2} = (a(A \cap b))^{\alpha_2} = a(A \cap b)$ since $\alpha_1 \in a(A \cap b)$ for the first step, and $a \lhd A, b \lhd B$ for the second step.

Now $a(A \cap B)/a(A \cap b) = (A \cap B)a(A \cap b)/a(A \cap b)$ since $a(A \cap B) = (A \cap B)a(A \cap b)$. To see this, note that if $E = (A \cap B), D = a(A \cap b)$ then we saw above that $E < N_G(D)$, and hence $ED = DE$. Thus $(A \cap B)a(A \cap b) = a(A \cap b)(A \cap B) = a(A \cap B)$ since $A \cap b < A \cap B$.

Now letting $H = A\cap B, K = a(A\cap b)$ then we have $a(A\cap b)/a(A\cap b) = HK/K \cong H/(H\cap K) = A\cap B/[(A\cap B)\cap(a(A\cap B))]$ by the Second Isomorphism Theorem 1.12.

Then we will show $A\cap B/[(A\cap B)\cap(a(A\cap b))] = A\cap B/a(A\cap b)\cap(a\cap B)b$. The latter is symmetric, so we're done once we show this. So we need to show $(A\cap B)\cap(a(A\cap b)) = a(A\cap b)\cap(a\cap B)b$. To see it, we show both inclusions. Pick $x\in(A\cap B)\cap(a(A\cap b))$. Then $x\in a(A\cap b)$. Write $x = \alpha\beta$ where $\alpha\in a$ and $\beta\in A\cap b$. Then also $\alpha\beta\in A\cap B$. As $\beta\in b < B$, and $\alpha\beta\in B$, it follows that $\alpha\in B$. Then $\alpha\in a\cap B$, so indeed $\alpha\beta\in(a\cap B)b$. Thus $(A\cap B)\cap(a(A\cap b))\subseteq a(A\cap b)\cap(a\cap B)b$. Now take $x\in a(A\cap b)\cap(a\cap B)b$. Then $x\in a(A\cap b)$. Also, $x\in a(A\cap b)\subseteq A$, and $x\in(a\cap B)b\subseteq B$, so we're done. □

Exercise 1.23. Vector spaces are, in particular, additive groups. Use the same tricks as in the proof of Jordan-Hölder Theorem (including Butterfly Lemma) to show that any two bases of a finite-dimensional vector space V are equivalent, in the sense that they have the same cardinality.

The Jordan-Hölder Theorem told us that finite groups can be understood in terms of simple groups. We will now produce some examples of simple groups.

Definition 1.24. A group G is called **cyclic** if there is some $g\in G$ such that $G = \left\{g^k : k\in\mathbb{Z}\right\}$.

Suppose $G = \langle g\rangle$ is cyclic. If $g^n \neq e$ for all $n\in\mathbb{Z}$ then $g^k\neq g^l$ for all $k\neq l$, and the map $g^k\mapsto k$ witnesses $(G,\cdot)\cong(\mathbb{Z},+)$. Otherwise, let $n\in\mathbb{Z}^+$ be the smallest positive integer such that $g^n = e$. Then $g^0 = e, g, g^2, \dots, g^{n-1}$ are all distinct, and $G = \left\{e, g, g^2, \dots, g^{n-1}\right\}$. The map $g^k\mapsto k$ is then an isomorphism witnessing $(G,\cdot)\cong(\mathbb{Z}/n\mathbb{Z},+)$.

Claim 1.25. A cyclic group G is simple if and only if $G\cong\mathbb{Z}/p\mathbb{Z}$ for some prime p.

Proof. If G is infinite then $G\cong\mathbb{Z}$, and as $\mathbb{Z}$ is abelian, $2\mathbb{Z}$ is a non-trivial normal subgroup. So it suffices to show that $\mathbb{Z}/n\mathbb{Z}$ is simple if and only if n is prime.

Suppose that $n = p$ is prime. Then $|\mathbb{Z}/p\mathbb{Z}| = p$. If $N < \mathbb{Z}/p\mathbb{Z}$ then we have $|N| \mid p$, so $\mathbb{Z}/p\mathbb{Z}$ has no non-trivial subgroups at all, and hence no non-trivial normal subgroups, so it is simple.

On the other hand, suppose n is not prime. Let p be a prime divisor of n. Consider $0, p, 2p, 3p, \dots, \frac{n}{p}p = n = 0$ in $\mathbb{Z}/n\mathbb{Z}$. This set is all distinct, and is clearly a non-trivial subgroup. As $\mathbb{Z}/n\mathbb{Z}$ is abelian it is also normal, so $\mathbb{Z}/n\mathbb{Z}$ is not simple. □

We will see some more examples of simple groups after a discussion of some properties of symmetric groups.

1.3. **Symmetric Groups.** There is a group homomorphism sgn : $S_n\to\{\pm1\}$, where we view $\{\pm1\}$ as a group under multiplication. The map is $\sigma\mapsto\operatorname{sgn}(\sigma) =: (-1)^\sigma =: (-)^\sigma$. The informal definition is as follows: A permutation in S_n is a way of reordering n numbers. If we represent σ by drawing $1, 2, 3, \dots, n$ above $1, 2, 3, \dots, n$ and then draw arrows from i on the top to $\sigma(i)$ on the bottom (in as generic a way as possible, so there at most two curves cross at any point, and no curves meet tangently), then $\operatorname{sgn}(\sigma)$ is $(-1)^{\text{number of crossings}}$. We claim that this is well-defined. To see this, one thinks topologically (insert picture). Also, we should note that this is a group morphism. To see this, one draws two pictures, then makes the composition by attaching i from the bottom row of the first picture to i of the top row of the second picture. The resulting picture is the picture for the composition, so one gets the total number of crossings by adding the numbers from the two pictures. (Insert picture). Now we work more formally:

Definition 1.26. $(-1)^\sigma := \prod_{i<j} \operatorname{sgn}(\sigma(j) - \sigma(i))$, where $\operatorname{sgn}(\sigma(i))$ is the sign of $\sigma(i)$ as a natural number.

Note that the above definition is clearly well-defined.

Claim 1.27. $(-1)^{\sigma\tau} = (-1)^\sigma(-1)^\tau$

Proof.

$$\begin{aligned}
(-1)^{\sigma\tau} &= \prod_{i<j} \operatorname{sgn}(\sigma\tau(j) - \sigma\tau(j)) \\
&= \prod_{i<j} \operatorname{sgn}(\tau(j) - \tau(i)) \cdot \left(\prod_{i<j} \frac{\operatorname{sgn}(\sigma\tau(j) - \sigma\tau(i))}{\operatorname{sgn}(\tau(j) - \tau(i))} \right) \\
&= (-1)^\tau \prod_{k \neq l, \text{each pair taken once, } k=\tau(i), l=\tau(j)} \frac{\operatorname{sgn}(\sigma(l) - \sigma(k))}{\operatorname{sgn}(l-k)} \\
&= (-1)^\tau \prod_{k<l} \frac{\operatorname{sgn}(\sigma(l) - \sigma(k))}{\operatorname{sgn}(l-k)} \\
&= (-1)^\tau(-1)^\sigma \\
&= (-1)^\sigma(-1)^\tau
\end{aligned}$$

□

Definition 1.28. A permutation σ such that $\operatorname{sgn}(\sigma) = 1$ is called **even**. When $\operatorname{sgn}(\sigma) = -1$ we call σ **odd**.

Definition 1.29. The **alternating group** is $A_n = \ker \operatorname{sgn}$, which is the collection of all even permutations.

Definition 1.30. A **transposition** is a permutation σ such that σ interchanges two consecutive values, and leaves all others fixed.

Claim 1.31. Every $\sigma \in S_n$ can be written as a product of transpositions.

Proof. Easy. □

Corollary 1.32. *A_n is those permutations which can be written as an even number of transpositions.*

Proof. Easy, since any transposition is odd. □

Definition 1.33. A **2-cycle** is a permutation which interchanges two values, and leaves all others fixed. If σ switches k and l, we write $\sigma = (kl)$.

Claim 1.34. $(-1)^{(kl)} = -1$.

Proof. Easy. □

Corollary 1.35. *A_n is the collection of all permutations that are a product of an even number of 2-cycles.*

Proof. Immediate from the claim. □

Note that we can see immediately that $|A_n| = \frac{1}{2}|S_n| = \frac{n!}{2}$.

We will eventually show:

Theorem 1.36. *A_n is simple for all $n \neq 1, 2, 4$.*

We note immediately that for $n = 1, 2$ we have $|A_n| = 1$, so A_n is not simple. For $n = 3$, we have $|A_3| = \frac{4!}{2} = 3$, and the only group of three elements is $\mathbb{Z}/3\mathbb{Z}$, which we have seen before is simple. So only the case $n \geq 4$ remains. Before we can handle this more difficult case, we discuss cycle decompositions for permutations.

Consider the permutation $[2, 1, 4, 5, 3]$. We can write this alternatively in cycle notation as $(12)(345)$. We write 1, then $\sigma(1)$, then $\sigma(\sigma(1))$, and so on until we reach 1 again. Then we close the parentheses, and start another one with the least element not yet written. For another example, $(14)(532)$ represents the same permutation as $[4, 5, 2, 1, 3]$. We call each parenthesized expression a **cycle**, and call it an n**-cycle** if it has n entries. Note that disjoint cycles permute, so $(12)(345) = (345)(12)$. Also, we may cyclically permute elements within a cycle, so $(345) = (534)$. Note that an n-cycle is even if and only if n is odd.

Claim 1.37. Every permutation can be written as a product of disjoint cycles. This decomposition is unique up to the ordering of the cycles and cyclic permutations within the cycles.

Claim 1.38. $(-1)^\sigma = (-1)^{\text{number of even-length cycles in its decomposition}}$.

Claim 1.39. Suppose $\sigma = (a_1 \dots a_k)$ (the following will be true also if σ is written as a product of disjoint cycles, but we do not do so for convenience). Suppose $\tau = [\tau(1)\tau(2)\dots\tau(n)]$. Then $\sigma^\tau = \tau^{-1}\sigma\tau = (\tau^{-1}(a_1)\tau^{-1}(a_2)\dots\tau^{-1}(a_k))$.

Proof. Let $L = \sigma^\tau$, and $R = (\tau^{-1}(a_1)\dots\tau^{-1}(a_k))$. Then $R(\tau^{-1}(a_1)) = \tau^{-1}(a_2)$. And $L(\tau^{-1}(a_1)) = \tau^{-1}(a_2)$ as well. All other $\tau^{-1}(a_i)$'s are equally clear. Also, it is clear that if $j \neq \tau^{-1}(a_i)$ for any i, then $L(j) = R(j) = j$. □

Definition 1.40. We say that $g_1, g_2 \in G$ are **conjugate** if there is $g \in G$ such that $g_1^g = g_2$.

Claim 1.41. Conjugacy is an equivalence relation on G.

Proof. Suppose $g_2^g = g_1$ and $g_3^{g'} = g_2$. Then $g_3^{g'g} = g_1$. This verifies transitivity. The other properties are even easier. □

By the above, G is subdivided into equivalence classes, which we call **conjugacy classes**.

Example 1.42. Take $G = S_n$. Then $\sigma_1, \sigma_2 \in S_n$ are conjugate if and only if they have the same list of cycles lengths. For example, $(453)(12) \in S_5$ is conjugate to any permutation whose disjoint cycle decomposition has a 3-cycle and a 2-cycle.

To see this, use Claim 1.39.

Corollary 1.43. *The number of conjugacy classes of S_n is equal to the number of ways to write n as a sum of positive integers, without regard to order. This latter number is called P_n, the number of partitions of n.*

Example 1.44. We consider S_3. We can write $3 = 3 = 2 + 1 = 1 + 1 + 1$, and there are no other ways of partitioning 3, so there are 3 conjugacy classes in S_3. One conjugacy class is $\{(123), (132)\}$. Another consists of all 2-cycles, which is $\{(12), (23), (31)\}$. Finally, another conjugacy class has only 1-cycles, so is $\{I\}$. Counting reveals that these are all the elements of S_3. Also, the above descriptions of A_n let us see immediately that $A_3 = \{I, (123), (132)\}$.

We are almost ready to prove Theorem 1.36. Before doing so, we consider the case $n = 4$. Recall Example 1.2(4), which gave a map $\phi : S_4 \to S_3$. We consider $\ker \phi$. By inspection, we see that $\ker \phi = \{I, (12)(34), (13)(24), (14)(23)\} \trianglelefteq A_4$. Also note $\ker \phi \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. So we see that A_4 is not simple. Note that this group also gives a counterexample to the claim that normality of subgroups is transitive. We have $N = \{I, (12)(34)\} \trianglelefteq \ker \phi \trianglelefteq S_4$, but $N \ntrianglelefteq S_4$. Now let us work toward showing that A_n is simple for $n > 4$.

Lemma 1.45. *Every element of A_n is a product of (not necessarily disjoint) cycles of length* 3*.*

Proof. If $\sigma \in A_n$ then we can write σ as a product of an even number of (not necessarily disjoint) 2-cycles. So it suffices to show that a product of two 2-cycles is always equal to a product of 3-cycles. There are two cases to consider, namely when the two 2-cycles are disjoint and when they are not.

In the not disjoint case we have, for example, $(23)(12) = (132)$, and there is clearly nothing special about the roles of $1, 2, 3$ in this example. In the case when they are disjoint, [see picture]. □

Lemma 1.46. *If $N \trianglelefteq A_n$ contains a* 3*-cycle, then $N = A_n$.*

Proof. It suffices to show that N contains every 3-cycle, then apply Lemma 1.45. By relabelling if necessary, assume $(123) \in N$. Given another 3-cycle (ijk), we know that $(ijk) = (123)^\sigma$ for some $\sigma \in S_n$. If $\sigma \in A_n$, then we're done, since $N \trianglelefteq A_n$. Otherwise, write $\sigma = (12)\sigma'$, so $\sigma' = (12)\sigma$. Now we have $\sigma \notin A_n$, so σ was odd. Thus σ' is even, so $\sigma' \in A_n$. Now we have $(123)^{(12)\sigma'} = (132)^{\sigma'} = ((123)^2)^{\sigma'}$. Now $(123) \in N$, so $(123)^2 \in A_n$, so we get $(ijk) = (123)^{(12)\sigma'} \in N$ as before. □

We are now ready to show that A_n is simple for $n > 4$.

Proof of Theorem 1.36. Suppose that $n > 4$, and $\{e\} \neq N \trianglelefteq A_n$. By Lemma 1.46 we are done if we can show that N contains a 3-cycle.

First, suppose there is some $\sigma \in N$ whose cycle decomposition contains a cycle of length ≥ 4. For concreteness, suppose $\sigma = (12345)m$ where m is the rest of the cycle decomposition (so consists of products of more disjoint cycles). Consider $\sigma^{(123)}$ (any cycle of length 3 will do in place of (123)). We get

$$\begin{aligned}\sigma^{-1}\sigma^{(123)} &= \sigma^{-1}(132)\sigma(123)\\ &= m^{-1}(12345)^{-1}(132)(12345)m(123)\\ &= (235)\end{aligned}$$

Now $\sigma^{-1} \in N$ and $\sigma^{(123)} \in N$, so we got $(235) \in N$, and we're done in this case.

We now see that we're done if there is a cycle of length ≥ 4. So suppose that every cycle has length at most 3.

Suppose there is some $\sigma \in N$ such that $\sigma = (123)(456)m$ (again, m is just the rest of the cycle decomposition). We consider $\sigma^{-1}\sigma^{(124)}$ (any 3-cycle involving something from each of (123) and (124) will do). Again we get that this is in N, and we have

$$\begin{aligned}\sigma^{-1}\sigma^{(124)} &= \sigma^{-1}(142)\sigma(124)\\ &= (12435\ldots)\end{aligned}$$

So we found a cycle of length ≥ 4 in N, so by the first case above we get $N = A_n$.

For the next case, suppose that N has an element of the form $\sigma = (123)m$ where m has no length ≥ 3 cycles, so m is a product of disjoint 2-cycles. Then $m^2 = I$, so $\sigma^2 = (132) \in N$. So N contains a 3-cycle, so by Lemma 1.46 we get $N = A_n$.

We are now reduced to the case where N consists only of elements which are products of disjoint 2-cycles. We have some $\sigma = (12)(34)m \in N$ where, as before, m is the rest of the cycle decomposition of σ. We have $\sigma^{-1}\sigma^{(123)} \in N$ (again, any 3-cycle meeting two distinct 2-cycles in σ will do), and we compute:

$$\begin{aligned}\sigma^{-1}\sigma^{(123)} &= \sigma(123)^{-1}\sigma(123)\\ &= (14)(23)\end{aligned}$$

So $\sigma' = (14)(23) \in N$. Since $n > 4$ we can now bring in the element 5, and get $\sigma'^{-1}\sigma'^{(125)} \in N$.

$$\begin{aligned}\sigma'^{-1}\sigma'^{(125)} &= \sigma'^{-1}(152)\sigma'(125)\\ &= (12345)\end{aligned}$$

So we again found a cycle of length ≥ 4, so by the first case we again get $N = A_n$. As we have exhausted all the cases, we're done. □

1.4. **Groups Acting on Sets.**

Definition 1.47. Let G be a group, X a set. We define G **acting on** X (or **a** G **action on** X, or X **is a** G**-set**) to by a homomorphism $\phi : G \to S(X)$, where $S(X)$ is the group of bijections on X. This gives rise to an **action**, which is a binary operation $* : G \times X \to X$ given by $g * x = \phi(g)(x)$.

Alternatively, we can begin with a binary operation $* : G \times X \to X$ satisfying $(g_1g_2) * x = g_1 * (g_2 * x)$ and $e * x = x$. Such a $*$ gives rise to a $\phi : G \to S(X)$, namely $\phi(g)(x) = g * x$.

Example 1.48.
1. G acts on itself by multiplication. Here $X = G$, and $g_1 * g = g_1g$.
2. G acts on itself by conjugation. Here again $X = G$, and $g_1 * g = g^{g_1^{-1}} = g_1gg_1^{-1}$.

   To see that this is an action, we check $e * x = x^{e^{-1}} = x$, and

$$\begin{aligned}(g_1g_2) * x &= x^{(g_1g_2)^{-1}}\\ &= x^{g_2^{-1}g_1^{-1}}\\ &= (x^{g_2^{-1}})^{g_1^{-1}}\\ &= g_1 * (g_2 * x)\end{aligned}$$

(3) This example refers to the Symmetry Gallery on Prof. Bar-Natan's website. Each row represents an action of some $G_i (i = 1, \ldots, 17)$ on $\mathbb{R}^2$.

For example, any tiling of the plane has symmetries that come from moving the grid up n units or left n units, for any $n \in \mathbb{Z}$. Thus $\mathbb{Z}^2$ acts on a plane tiling. This is the last row of the symmetry gallery.

Consider the bricklaying pattern. If G is the group of symmetries of this picture, then $G \supsetneq \mathbb{Z}^2$, and G acts on $\mathbb{R}^2$.

It is true, but we will not show it, that (up to some conditions) there are exactly 17 tiling patters (i.e., group actions on $\mathbb{R}^2$, and they are exactly the ones in the Symmetry Gallery.

(4) If $H \leq G$, then G/H may not be a group, but it is a set. Let $X = G/H$. Here G acts by $g' * [g] = g' * gH = (g'g)H = [g'g]$. This action is **transitive**, meaning that for any $x, y \in X$, there is some $g \in G$ such that $g * x = y$.

(5) If X_1, X_2 are G-sets, then their disjoint union $X_1 \sqcup X_2$ is also a G-set in a natural way, using the action of G on X_1 for elements in X_1, and the action of G on X_2 for elements of X_2. In this case, the action cannot be transitive unless $X_1 = \emptyset$ or $X_2 = \emptyset$.

Fact 1.49. *Fix a group G. Then the collection of G-sets forms a category. The objects are G-sets. For the morphisms, given X_1, X_2 both G-sets, a **morphism of G-sets** is a map $f : X_1 \to X_2$ such that for every $g \in G$ and every $x_1 \in X_1$, $g * f(x_1) = f(g * x_1)$. That is, for any $g \in G$, the following diagram commutes: An **isomorphism of G-sets** is an invertible morphism of G-sets.*

As an aside, in topology one has the following theorem, which says that G-sets can be useful in understanding topological spaces:

Theorem 1.50. *Given a well-behaved "base space" B, the collection of all coverings of B is a category, and this category is naturally equivalent to the category of G-sets, where $G = \Pi_1(B)$, the fundamental group of B.*

Now back to the algebra:

Theorem 1.51. *Fix a group G.*

(1) Every G-set is a (possibly infinite) disjoint union of transitive G-sets.

(2) Every transitive G-set is isomorphic to a G-set of the form G/H for some $H \leq G$.

Proof. (1) Given $x \in X$, the **orbit of** x is $\text{orb}_G(x) = \{g * x : g \in G\}$. Define a relation $\sim$ on X by $x_1 \sim x_2 \iff \exists g \in G, x_2 = g * x_1$. Then the orbit of x is the the equivalence class of x under $\sim$. Then it is clear that X is a disjoint union of orbits, and every orbit, considered in itself, is a transitive G-set.

(2)

□

Theorem 1.52. *if G is a p-group then $Z(G)$ is non-trivial.*

Proof. Let G act on itself by conjugation. Let the x_i's be representatives from the non-trivial conjugacy classes (i.e., orbits) of G. Then we have, since $x \in Z(G)$ if and only if x is an orbit of size 1,

$$\begin{aligned}|G| &= |Z(G)| + \sum_i |O(x_i)| \\ &= |Z(G)| + \sum_i [G : \mathrm{Stab}_G(x_i)] \\ &= |Z(G)| + \sum_i [G : C_G(x_i)]\end{aligned}$$

Here $C_G(x_i) = \{y \in G : x_i y = y x_i\} \leq G$ is the centralizer of x_i in G. This equation, which is called the **class equation**, implies the theorem. Indeed, $|G|$ and $\sum_i [G : C_G(x_i)]$ are divisible by p. Thus $|Z(G)|$ is divisible by p, and in particular is not 1. □

Definition 1.53. Given a finite group G, a **Sylow p-subgroup** is a subgroup $P \leq G$ such that $|P| = p^\alpha$ for some α, and such that $p^\alpha \mid |G|$, and $p^{\alpha+1} \not| |G|$. That is, P is a maximal p-subgroup of G. We let $\mathrm{Syl}_p(G)$ be the set of Sylow p-subgroups of G.

Theorem 1.54 (Sylow Theorems)**.** *Given a finite group G,*

(1) $\left|\mathrm{Syl}_p(G)\right| \equiv 1 \mod p$.
(2) Every p-subgroup of G is contained in a Sylow p-subgroup.
(3) All Sylow p-subgroups of G are conjugate.

Corollary 1.55. $\left|\mathrm{Syl}_p(G)\right| \mid |G|$.

Proof. By the theorem all such groups are conjugate. Letting G act by conjugation on $\mathrm{Syl}_p(G)$, we see that this action is transitive, and the result follows. □

We will prove this later, but for now we give a few examples of its application.

Example 1.56. We find all groups of order 15.

We know that $|\mathrm{Syl}_5(G)| \equiv 1 \mod 5$. On the other hand, $|\mathrm{Syl}_5(G)| \mid |G| = 15$. Thus $|\mathrm{Syl}_5(G)| = 1$. Let $P_5 \leq G$ by the Sylow 5-subgroup of G. Observe:

Corollary 1.57. *If P is the unique Sylow p-subgroup of G, then $P \trianglelefteq G$.*

Proof. All conjugates of P are Sylow p-subgroups, and hence are P. Thus P is closed under conjugation, so is normal. □

So we have $P_5 \trianglelefteq G$. On the other hand, we also get that there is a unique Sylow 3-subgroup, call it $P_3 \trianglelefteq G$.

Notice also, more generally, that if $|G| = pq$ for distinct primes $p < q$, then G has a unique Sylow q-subgroup. If also $p \nmid q - 1$ then we also get a unique Sylow p-subgroup. Everything we are doing in this example works equally well in this context.

Here is an aside, which we could have proved much earlier:

Proposition 1.58. *If H is a group such that $|H| = p$ for some prime p, then $H \cong \mathbb{Z}/p\mathbb{Z}$.*

Proof. Pick some $x \in H$ such that $x \neq e$. Consider $\langle x \rangle \leq H$. Since $x \neq H$ we have $\langle x \rangle \neq \{e\}$. Also $|\langle x \rangle| \mid |H|$, so as H is of prime order $\langle x \rangle = H$. The isomorphism we need is a map $\phi : \mathbb{Z}/p\mathbb{Z} \to H$ given by $\phi(a) = x^a$. It is easy to check that this is an isomorphism. □

In our case, we have $P_5 \cong \mathbb{Z}/5\mathbb{Z}$ and $P_3 \cong \mathbb{Z}/3\mathbb{Z}$. Consider any $y \in P_3$. We claim that y commutes with all elements of P_5. To see this, note that if $x \in P_5$, then since P_5 is normal we have $x^y \in P_5$. Thus the map $x \mapsto x^y$ is an automorphism of P_5. Now we apply the next result:

Proposition 1.59. $\mathrm{Aut}(\mathbb{Z}/p\mathbb{Z}) \cong (\mathbb{Z}/p\mathbb{Z})^*$ *for any prime* p.

Proof. Let ϕ be an automorphism of $\mathbb{Z}/p\mathbb{Z}$. Since $\mathbb{Z}/p\mathbb{Z}$ is cyclic, ϕ is determined by its value on 1. Since ϕ is surjective, one easily checks that we can have $\phi(1)$ any value except 0. We thus get a bijection $\phi \mapsto \phi(1)$, and it is easy to check that this is an isomorphism of groups. □

We have $|(\mathbb{Z}/p\mathbb{Z})^*| = p-1$, and we will show later that $(\mathbb{Z}/p\mathbb{Z})^* \cong \mathbb{Z}/(p-1)\mathbb{Z}$.

Back to our situation, we can apply this result to see that $|\mathrm{Aut}(P_5)| = 4$. Thus the order of the map ϕ given by $\phi(x) = x^y$, has $|\phi| \mid 4$. But also $|\phi| \mid |y| \mid 3$. This implies that $|\phi| = 1$, so $x = x^y$, so $xy = yx$. So indeed y commutes with all elements of P_5.

Also, $P_3 \cap P_5 = \{e\}$. This is because $|P_3 \cap P_5| \mid 3$ and $|P_3 \cap P_5| \mid 5$. Consider P_3P_5. By the Second Isomorphism Theorem this is a group of order 15. Thus $P_3P_5 = G$.

Proposition 1.60. *If* $G = G_1G_2$ *with* $G_1 \cap G_2 = \{e\}$ *and such that* $[G_1, G_2] = \{e\}$ *(i.e., all elements of* G_1 *commute with all elements of* G_2*), then* $G \cong G_1 \times G_2$.

Proof. Define $\phi : G \to G_1 \times G_2$ by $\phi(g_1g_2) = (g_1, g_2)$. One needs to check that this is well-defined. If $g_1g_2 = g_1'g_2'$ then $g_1{g_1'}^{-1} = g_2'g_2^{-1}$. The left side is in G_1, the right in G_2, and $G_1 \cap G_2 = \{e\}$, so $g_1{g_1'}^{-1} = e$, and so $g_1 = g_1'$. Similarly, $g_2 = g_2'$. So ϕ is well-defined. It is then easy to check that ϕ is an isomorphism. □

Back to our example, we have shown that $G \cong \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z}$. We need one more thing:

Proposition 1.61. *Let* m, n *be relatively prime integers. Then* $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}/(mn)\mathbb{Z}$.

Proof. This is a version of the Chinese Remainder Theorem. □

Thus we see, in our example, that $G \cong \mathbb{Z}/15\mathbb{Z}$. So there is exactly one group of order 15, which is the cyclic group of order 15.

The above example shows, in fact, that the only group of order pq where $p < q$, $p \nmid q-1$ is $\mathbb{Z}/(pq)\mathbb{Z}$. On the other hand, if we look at groups of order 21, we get that P_3 might not be normal. Moreover, the argument that elements of P_3 commute with elements of P_7 does not apply. If we get that a conjugation automorphism has order 3 (instead of 1), then $[P_3, P_7] \neq \{e\}$. We thus do not get the direct product. Instead, we get a semidirect product, which we will see in more detail later.

We now return and prove Theorem 1.54. We will need the following tools:

Proposition 1.62. *Every* G*-set is a union of orbits, each of order dividing* $|G|$.

Proof. We have seen this before. □

Proposition 1.63 (Cauchy's Theorem)**.** *If* G *is an abelian group with* $p \mid |G|$ *with* p *a prime, then* G *has an element of order* p.

Proof. Induction on $|G|$. Pick any $x \in G$, and consider $\langle x \rangle$. If $p \mid |x|$ then $x^{|x|/p}$ is an element of order p. Otherwise, $p \mid |G| / |\langle x \rangle| = |G/\langle x \rangle|$. By induction there is some $y \in G$ such that $[y]$ has order p in $G/\langle x \rangle$. Then $y^p \in \langle x \rangle$, so $y^p = x^\alpha$ for some α. Write $|y| = kp + r$ with $0 \leq r < p$. Then $e = y^{|y|} = y^{kp}y^r = x^{\alpha k}y^r$. Now $x^{\alpha k} \in \langle x \rangle$, so $y^r \in \langle x \rangle$. That is, $[y]^r = e_{G/\langle x \rangle}$. But $0 \leq r < p = |[y]|$, so $r = 0$. So $|y| = kp$, and y^k is of order p. □

Proposition 1.64. *Fix $P \in \mathrm{Syl}_p(G)$. If $x \in G$ is such that $|x| = p^b$ (or $H < G$ is a p-group) and x normalizes P (or $H \leq N_G(P)$), then $x \in P$ (or $H \leq P$).*

Proof. It suffices to do the case for the p-group H, since if x is a p-element then $\langle x \rangle$ is a p-group.

By the Second Isomorphism Theorem, since $H \leq N_G(P)$, $HP \leq G$, and $|HP| = \frac{|H||P|}{|H \cap P|}$. By our hypotheses we get that $|HP|$ is a power of p. Moreover, $HP \supseteq P$, and P a Sylow p-group, so $|HP| = |P|$. Thus $|H| = |H \cap P|$, so $H \subseteq P$. □

Proof of Theorem 1.54. We recall $p^\alpha \mid |G|$ and $p^{\alpha+1} \nmid |G|$.

Claim 1.65. $\mathrm{Syl}_p(G) \neq \emptyset$.

Proof. First, note that if $p \nmid |G|$ then there is nothing to do, the Sylow p-group will be $\{e\}$. Now we go by induction on $|G|$. We use the class equation

$$|G| = |Z(G)| + \sum_i \frac{|G|}{|C_G(x_i)|}$$

As usual the x_i's are representatives of the non-trivial conjugacy classes of G. We have $p \mid |G|$, so either $p \mid |Z(G)|$, or there is some i such that $p^\alpha \mid |C_G(x_i)|$.

In the latter case, $C_G(x_i) \lneq G$. By induction, $C_G(x_i)$ has a Sylow p-subgroup P. So $P \leq G$. But $p^\alpha \mid |C_G(x_i)|$, so $|P| = p^\alpha$, and P is a Sylow p-subgroup of G.

In the former case, by Proposition 1.63, $Z(G)$ contains some N such that $|N| = p$, and $N \trianglelefteq G$ since $N \leq Z(G)$. Then $|G/N| = |G|/p$. By induction, let $H \leq G/N$ be a Sylow p-subgroup of G/N. Then $|H| = p^{\alpha-1}$. Let $P = \pi^{-1}(H)$. Then $|P| = |H|\, p = p^\alpha$, so P is the p-group we needed. □

Claim 1.66. If $P \in \mathrm{Syl}_p(G)$, letting X be the set of conjugates of P by elements of G, we have $|X| \equiv 1 \mod p$.

Proof. P acts on X by conjugation. At least one orbit is a singleton, namely the orbit containing P. We will have the claim if we can show that every other orbit has size divisible by p. Suppose that $P \neq P'$, and $P' \in X$. $|\text{orbit of } P'| = |P| / |\mathrm{Stab}_P(P')|$. We see that this is always a power of p, and must see that it is never $p^0 = 1$.

To see this, note that $|P'| = |P| = p^\alpha$ since P' is a conjugate of P, so $|P'|$ is a Sylow p-subgroup of G. If $|\mathrm{orbit}(P')| = 1$ then every element of P acts on P' trivially, so $P \leq N_G(P')$. By Proposition 1.64, it follows that $P \subseteq P'$. But they have the same order, so $P = P'$, contradicting our choice of P'. □

Now we fix a Sylow p-subgroup of G, say P.

Claim 1.67. If $H \leq G$ is a p-group, then H is contained in a conjugate of P.

Observe that proving this claim suffices to prove the theorem.

Proof. Let X_P be the set of conjugates of P by elements of G. Then H acts on X_P. Note that $p \nmid |X_P|$. Every orbit has size divisible by p (since H is a p-group), so there must exist some $P' \in X_P$ which is a singleton orbit. This means that $H \leq N_G(P')$. So by Proposition 1.64 $H \subseteq P'$. $\square$

$\square$

1.5. **Semidirect Products.** Fix a group G. We do not require G to be finite. Suppose that $H, N \leq G$. Consider the function $\mu : H \times N \to HN$ given by $(h, n) \mapsto hn$. Recall that, in general, HN is not a group, so this is not a morphism. In general, μ is not injective. Indeed, if $H \cap N \neq \{e\}$ then μ is not injective. However,

Claim 1.68. If $H \cap N = \{e\}$, then μ is injective.

Proof. Suppose $h_1n_1 = h_2n_2$. Then $n_1n_2^{-1} = h_1^{-1}h_2$ is in $H \cap N = \{e\}$, so $n_1 = n_2$ and $h_1 = h_2$. $\square$

Consider the example where $G = S_3$, $H = \langle(1, 2)\rangle$, and $N = \langle(2, 3)\rangle$. Then $\langle HN\rangle = G$, so $|\langle HN\rangle| = 6$, but $|HN| = 4$, so HN is not a group. From here on we assume that $H \cap N = \{e\}$.

Now suppose that $H \trianglelefteq G$ and $N \trianglelefteq G$. Then μ is an isomorphism. Indeed, $\mu((h_1, n_1)(h_2, n_2)) = \mu(h_1h_2, n_1n_2) = h_1h_2n_1n_2$. On the other hand, $\mu(h_1, n_1)\mu(h_2, n_2) = h_1n_1h_2n_2$. So to see that μ is a morphism we must show that $[H, N] = \{e\}$. To see this, take any $h \in H, n \in N$. Then $hnh^{-1}n^{-1} = n^{h^{-1}}n^{-1} \in N$ since N is normal. But also $hnh{-}1n^{-1} = h(h^{-1})^{n^{-1}} \in H$, since H is normal. Since $N \cap H = \{e\}$, we see that $hnh^{-1}n^{-1} = e$, so indeed $[H, N] = \{e\}$.

From the above, we see that the interesting case is when one of the subgroups is normal, but the other is not. This is the case we will describe in what follows. So we work in the case $N \trianglelefteq G, H \leq G, H \cap N = \{e\}$. From the above we see that $\mu : H \times N \to HN$ is a bijection, but not generally an isomorphism. Our goal is to understand the group structure of HN in terms of the groups structures on H and N.

Notice that in this case H acts on N by conjugation, $h \mapsto (n \mapsto n^h)$. Let $\phi_h : N \to N$ denote the conjugation by h maps. Then $\phi_h \in \mathrm{Aut}(N)$, so the map $h \mapsto \phi_h$ is a group morphism from H to $\mathrm{Aut}(N)$. We would like to write $h_1n_2 \cdot h_2n_2$ as something in H times something in N, for we will then have described the product in HN. We use a familiar trick: $h_1n_1h_2n_2 = h_1h_2h_2^{-1}n_1h_2n_2 = h_1h_2n_1^{h_2}n_2$. Now $h_1h_2 \in H$, and $n_1^{h_2}n_2 \in N$. We rewrite this as $\mu(h_1h_2, \phi_{h_2}(n_1)n_2)$. So the product has the first n "twisted" by the second h. This inspires the following definition:

Definition 1.69. Given any two groups H, N, and a morphism $\phi : H \to \mathrm{Aut}(N)$, we define the **semi-direct product of** H **and** N by $H \ltimes_\phi N = \{(h, n) : h \in H, n \in N\}$, with multiplication given by $(h_1, n_1)(h_2, n_2) = (h_1h_2, \phi_{h_2}(n_1)n_2)$. We usually write $\ltimes$ for $\ltimes_\phi$ when ϕ is clear from context.

The above definition is incorrect (essentially because conjugation is an antimorphism, not a morphism). Here are some ways to fix it:

1. Could take ϕ to be an antimorphism. That is, $\phi(ab) = \phi(b)\phi(a)$.
2. Given any group G, consider the opposite group G^{op}, which as a set is G, but has operation $*$ where $a * b = ba$. It turns out that $G \cong G^{\mathrm{op}}$, by the

isomorphism $g \mapsto g^{-1}$. Then we could have taken ϕ to be a morphism $\phi : H^{\mathrm{op}} \to \mathrm{Aut}(N)$, or ϕ a morphism $\phi : H \to \mathrm{Aut}(N)^{\mathrm{op}}$.

(3) Instead of HN, look at NH. Then we would have obtained $n_1h_1n_2h_2 = n_1h_1n_2h_1^{-1}h_1h_2 = n_1n_2^{h_1^{-1}}h_1h_2 = n_1\psi_{h_1}(n_2)h_1h_2$. Now $\psi : H \to \mathrm{Aut}(N)$ is a genuine morphism.

(4) When we talked about G-sets, we said that we had a map $G \times X \to X$, or equivalently a morphism $G \to S_X$ such that $(g_1g_2) * x = g_1 * (g_2 * x)$. This should really be called a **left G-set**. Then a **right G-set** is a map $X \times G \to X$ such that $x * (g_1g_2) = (x * g_1) * g_2$. This is, equivalently, an antimorphism $G \to S_X$. The theory of left G-sets is the same as the theory of right G-sets, except that everything gets flipped. We could thus have switched languages in the definition of semidirect product, and taken a right H-action on N.

Strictly speaking, we should go back through the above and, every time we used G acting by conjugation, replace it by G acting on the right by conjugation (or on the left by conjugation by the inverse). But the conclusions remain unchanged.

Theorem 1.70. *(1) If $H \leq G, N \trianglelefteq G$, and $H \cap N = \{e\}$, then $H \ltimes_\phi N$, where ϕ is conjugation in G, has $H \ltimes N \cong HN$. The isomorphism is $\mu : H \ltimes N \to HN$, $\mu(h, n) = hn$.*

(2) In general,

(a) $H \leq H \ltimes N$, $N \trianglelefteq H \ltimes N$.

(b) $(H \ltimes N)/N \cong H$.

Proof. Part (1) is immediate from the construction. For part (2), we identify H with $\{(h, e_N) : h \in H\}$. It is easy to see that the product in $H \ltimes N$ restricted to H agrees with the product in H. Likewise, identify N with $\{(e_H, n) : n \in N\}$. Again, we see that the product in $H \ltimes N$ restricted to N agrees with the product in N. So $H \leq H \ltimes N, N \leq H \ltimes N$. To see that N is normal, we first observe that $(h, n)(h', n') = (e, e)$ implies $hh' = e$, and $\phi_{h'}(n)n' = e$. So $h' = h^{-1}$, and $n' = \phi_{h^{-1}}(n^{-1})$. Thus $(h, n)^{-1} = (h^{-1}, \phi_{h^{-1}}(n^{-1})$. Now we compute $(h_1, n_1)^{-1}(e_H, n)(h_1, n_1) = (h_1^{-1}, \phi_{h_1^{-1}}(n_1^{-1})(e_H, n)(h_1, n_1) = \ldots = (e_H, \phi_{h_1}(n)) \in N$.

For the second part of (2), the isomorphism is $(h, n)N \mapsto h$. □

Example 1.71. (1) $\{\pm 1\}$ acts on $\mathbb{Z}/n\mathbb{Z}$ by $\phi_1(k) = k$ and $\phi_{-1}(k) = -k$. We can thus form $\{\pm 1\} \ltimes \mathbb{Z}/n\mathbb{Z} =: D_{2n}$, the **dihedral group of order $2n$**. Geometrically, this group is the groups of symmetries (including flips, so not orientation-preserving) of a regular n-gon in the plane.

(2) Let $\mathbb{F}$ be a field. The linear functions with non-zero slope is $\{f(x) = ax + b : a, b \in \mathbb{F}, a \neq 0\}$. This is a group under composition. Two subgroups of this group are $\mathbb{F}_b^+ = \{x + b : b \in \mathbb{F}\}$ and $\mathbb{F}_a^\times = \{ax : a \in \mathbb{F}\}$. It is easy to check that $\{ax + b : a, b \in \mathbb{F}\} = \mathbb{F}_b^+ \ltimes \mathbb{F}_a^\times$, where the action is $\phi_a(b) = ab$, $\mathbb{F}^\times \to \mathrm{Aut}(\mathbb{F}^+)$.

(3) The above example can be generalized further. Let V be a vector space, and let $\mathrm{GL}(V)$ be the collection of all invertible linear transformations $V \to V$, as a group under composition. Consider $\{Ax + b : A \in \mathrm{GL}(V), b \in V\}$. This is the group of affine transformations of V. In the same way as above, $\{Ax + b : A \in \mathrm{GL}(V), b \in V\} = \mathrm{GL}(V) \ltimes V$. This generalizes the above, because $\mathbb{F}^\times = \mathrm{GL}(\mathbb{F})$.

(4) (a) Recall that SO(3) is the group of rotations in $\mathbb{R}^3$. The "symmetry group of physics before 1905" is $\mathrm{SO}(3) \ltimes \mathbb{R}^3$. It is also called the "Galilean group".
(b) Likewise, after Einstein, physics is better described by $\mathrm{SO}(3,1)$ (which is the "Lorentz group", we do not give a definition) and $\mathbb{R}^4$. That is, we have the "Poincaré group" $\mathrm{SO}(3,1) \ltimes \mathbb{R}^4$, the symmetry group of special relativity.

Example 1.72. In this example we describe the **Braid Group** B_n.

Consider

$$\begin{aligned}(\mathbb{C}^n \setminus \{\text{diagonals}\} &= \{(z_1, \dots, z_n) : \forall i \neq j, z_i \neq z_j\})/S_n \\ &= \{\text{polynomials of degree } n \text{ in } \mathbb{C}[X] \text{ with no repeated roots}\}.\end{aligned}$$

This is a topological group. Let B_n be its fundamental group.

As an alternative definition, start with {paths in $\mathbb{C}^n$}. Such a path appears like a braid, since the configuration of the points at $t = 0$ and $t = 1$ are the same, but we do not distinguish the points, so they may end at other places. Then B_n is that set of paths, modulo homotopy. That is, we allow the paths to be moved around, so long as we never cause an intersection. This is precisely what can be done with string, for example. In this presentation, the product of two paths is composition: follow the first path, then the other.

There is a third presentation, which we give now.

Definition 1.73. The **free group** on a set X is given as follows. We present in the case $X = \{a, b, c\}$, but the definition generalizes to any cardinality of X. The free group $\mathbb{F}(X)$ is the group containing X, also all the things forced by the definition of a group (inverses, products), but nothing else, and no additional relations. That is, it is $\left\{\text{words on } X \sqcup X^{-1}\right\}$, with multiplication being concatenation. If $|X| = n$, we often write $\mathbb{F}(n)$ for $\mathbb{F}(X)$.

Theorem 1.74. *This makes sense.*

The main problem in proving the theorem is that we need to impose $aa^{-1} = e = a^{-1}a$, where e is the empty word. To defeat this difficulty, we instead work with reduced words, where aa^{-1} never appear, with the product as concatenation followed by removal of aa^{-1} pairs. There is substantial bookkeeping required, but there is no conceptual difficulty.

The free group is universal, in the sense that if $i : X \to \mathbb{F}(X)$ is the inclusion map and $f : X \to G$ is a set map of X into some group G, then there is a unique group morphism $\phi : \mathbb{F}(X) \to G$ such that $\phi \circ i = f$. Clearly the map must be $\phi(abc) = \phi(a)\phi(b)\phi(c)$. This universal property characterizes $\mathbb{F}(X)$.

Definition 1.75. We write $\langle a, b, c : a^2 = b^4, cbc^{-1} = a\rangle$ (a **definition by generators and relations**), to mean the group $\mathbb{F}(\{a, b, c\})$/smallest normal subgroup containing $a^2b^{-4}, cbc^{-1}a^{-1}$.

We claim, but do not prove, that, if σ_i means that the first i things do not move, and i crosses over $i + 1$, then the rest do not move, then

$$B_n = \langle \sigma_1, \dots, \sigma_{n-1} : \sigma_i\sigma_j = \sigma_j\sigma_i \forall\, |i - j| > 1, \sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1} \forall i\rangle$$

The **pure braid group** is the subgroup of the braid group formed by **pure braids**, namely those braids which induce the identity permutation (that is, at the end of the braid, each strand is taken to the same position as it started at).

More precisely, comparing the starting and ending position of the strands induces a homomorphism $B_n \to S_n$, and the kernel is PB_n.

There is a map $\rho : PB_n \to PB_{n-1}$, given by dropping the last strand. It is true, but we will not prove it, that $\ker \rho = \mathbb{F}(n-1)$. The reason is that the strands sent to e can be described by a word in $\mathbb{F}(n-1)$ describing which strands it crosses over and under, once the first $n-1$ strands are straightened.

There is also an inclusion map $PB_{n-1} \to PB_n$ given by adding a straight strand on the right. Thus $PB_{n-1}, \mathbb{F}(n-1) \leq PB_n$. Clearly $PB_{n-1} \cap \mathbb{F}(n-1) = \emptyset$. As the kernel of a homomorphism, $\mathbb{F}(n-1) \trianglelefteq PB_n$. Thus $PB_n \cong PB_{n-1} \ltimes \mathbb{F}(n-1)$. Continuing by induction, $PB_n \cong (PB_{n-2} \ltimes \mathbb{F}(n-2)) \ltimes \mathbb{F}(n-1), \ldots$. At the end, we get $PB_n \cong ((\mathbb{F}(1) \ltimes \mathbb{F}(2)) \cdots) \ltimes \mathbb{F}(n-1)$.

To understand this, we start by understanding the action of PB_n on $\mathbb{F}(n)$. Given $\beta \in PB_n$, and $w \in \mathbb{F}(n)$, $\phi_\beta(w) = \beta^{-1} w \beta$, where we view w and β as being elements of PB_{n+1}, as above. See picture.

Here is another application of Sylow's Theorem:

Claim 1.76. Any group of order 12 is a semidirect product.

Proof. Suppose not, and let G be of order 12 which is not a semi-direct product. G has a Sylow 3-subgroup P_3 and a Sylow 2-subgroup P_2. So $|P_3| = 3$, and $|P_2| = 4$. Hence $P_2 \cap P_3 = \{e\}$. Also $P_3 \cong \mathbb{Z}/3\mathbb{Z}$ and it is easy to check that we have either $P_2 \cong \mathbb{Z}/4\mathbb{Z}$ or $P_2 \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. If one of these is normal then we know we get a semidirect product. So suppose that both are not normal. Then $n_2(G) \mid 12$, and $n_2(G) \equiv 1 \mod 2$, and $n_2(G) \neq 1$ since P_2 is not normal. So we get $n_2(G) = 3$. Similarly, we get $n_3(G) \mid 12$, $n_3(G) \equiv 1 \mod 3$, and $n_3(G) \neq 1$. So $n_3(G) = 4$. Since any two distinct subgroups of order 3 intersect trivially, we see that we have $2 * 4 = 8$ elements of order 3. Together with the identity, we have 9 elements, so only 3 are left. These 3 elements must all be in P_2. But this counts all the elements of G, and we no elements left for the conjugates of P_2. □

In light of the above, to understand all groups of order 12 it suffices to classify all relevant morphism. This is easy, and we find that if $|G| = 12$ then G must be one of:

1. $\mathbb{Z}/12\mathbb{Z}$
2. $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/6\mathbb{Z}$
3. A_4
4. D_{12}
5. $\mathbb{Z}/3\mathbb{Z} \rtimes \mathbb{Z}/4\mathbb{Z}$

For example, we can easily see that $\mathrm{Aut}(\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}) \cong S_3$. So if we are interested in morphism $\phi : \mathbb{Z}/3\mathbb{Z} \to \mathrm{Aut}(\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z})$ then these are the possibilities: We could have the trivial morphism, in which case we get a direct product $G = \mathbb{Z}/3\mathbb{Z} \times (\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}) \cong \mathbb{Z}/6\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. Otherwise, the generator of $\mathbb{Z}/3\mathbb{Z}$ must map to one of the cyclic permutations of order 3. In both cases we get A_4. The other examples are similar.

Definition 1.77. A group G is **solvable** if all of the factors G_i/G_{i+1} appearing in its decomposition series $G = G_0 \triangleright G_1 \triangleright G_2 \triangleright \cdots \triangleright G_n = \{e\}$ are abelian.

Note that we proved that the factors are uniquely determined up to permutation, so the above definition makes sense.

Example 1.78. A_4 is simple, but A_5 is not.

Theorem 1.79. *(1) If $N \trianglelefteq G$ then G is solvable if and only if both N and G/N are.*
(2) If $H \leq G$ and G is solvable, then H is solvable.

Proof. (1) Take the normal tower $G \triangleright N \triangleright \{e\}$, and refine it to a decomposition series for G. Then the result is immediate, using the fourth isomorphism theorem.

(2) If $B, A \leq G$ and $B \trianglelefteq A$ and A/B is abelian, then $B \cap H \trianglelefteq A \cap H$, and $A \cap H/B \cap H$ is also abelian. Assuming this, take a decomposition series of G, intersect it with H, and we get a decomposition series for H. That the quotients are abelian then also follows from the claimed statement.

To see the claim, note that it is clear that $B \cap H \trianglelefteq A \cap H$. Then $H \cap A/H \cap B \to A/B$ by the map $[a]_{H \cap B} \to [a]_B$. It is easy to see that this is a well-defined injective group morphism. So we can identify $H \cap A/H \cap B$ with a subgroup of A/B. As A/B is abelian, so is $H \cap A/H \cap B$.

□

2. Rings

2.1. **Basics.** For the remainder of the course by "ring" we will mean "commutative ring", unless explicitly stated otherwise.

Definition 2.1. A **field** is a ring F such that every non-zero element has a multiplicative inverse. A **domain** (or **integral domain**) is a ring with no zero-divisors. A **zero-divisor** is $x \in R$ such that $x \neq 0$ and there exists $y \neq 0$ such that $xy = 0$.

Many rings are constructed as quotients. We would like to know, given a ring R, for which ideals I is R/I a field or a domain.

Definition 2.2. Let R be a ring. An ideal I of R is called **maximal** if whenever J is an ideal of R and $J \supseteq I$ then either $J = I$ or $J = R$.

Definition 2.3. Let R be a ring, $S \subseteq R$. The **ideal generated by** S is the smallest ideal in R containing S. It is denoted by $\langle S \rangle$. If $S = \{x\}$ for some $x \in R$ we write $\langle x \rangle$ or xR for the ideal generated by S.

Proposition 2.4. *Let R be a commutative ring, $I \subseteq R$ an ideal. Then R/I is a field if and only if I is maximal.*

Proof. First, suppose that R/I is a field. Let J be an ideal such that $I \subsetneq J$. Then there is some $x \in J \setminus I$. Then $[x]_I \neq [0]_I$ in R/I, since $x \notin I$. So there is a multiplicative inverse to $[x]_I$, say $y \in R$ is such that $[x]_I[y]_I = [1]_I$. That is, $[xy]_I = [1]_I$. Thus $xy - 1 \in I \subsetneq J$. So $1 = xy - a$ for some $a \in J$. Since $x \in J, a \in J$ and J is an ideal, $xy - a \in J$. So $1 \in J$, so $J = R$. Thus I is maximal.

For the converse, suppose that I is maximal. Consider any $[x]_I \neq [0]_I$ in R/I. Then $x \notin I$. Consider $J = \langle I, x \rangle = I + \langle x \rangle = I + Rx$. Clearly $J \supsetneq I$ since $x \in J \setminus I$. Since I was maximal, $J = R$. So there is $a \in I, y \in R$ such that $a + yx = 1$. Thus in R/I, $[y]_I[x]_I = [1]_I$, so $[x]_I$ is invertible, and R/I is a field. □

Example 2.5. (1) Let p be a prime. Then $\langle p \rangle = p\mathbb{Z} \subseteq \mathbb{Z}$ is a maximal ideal. Indeed, $\mathbb{Z}/p\mathbb{Z}$ is a field.

(2) Consider the ring $l^\infty = \{\text{bounded sequences of real numbers}\}$. S is a ring under the pointwise operations inherited from $\mathbb{R}$. Its 0 element is $(0, 0, \ldots)$, and its 1 is $(1, 1, \ldots)$. Let $A_n = \{(a_i) \in l^\infty : a_n = 0\} \subseteq l^\infty$. It is easy to check that A_n is an ideal. We have $l^\infty/A_n \cong \mathbb{R}$. To see this, define $\pi_n : l^\infty \to \mathbb{R}$ by $(a_i) \mapsto a_n$. Then $\ker \pi_n = A_n$. Also, $\operatorname{im} \pi_n = \mathbb{R}$. It is easy to check that π_n is a ring homomorphism. So by the first isomorphism theorem $l^\infty/A_n \cong \mathbb{R}$. So l^∞/A_n is a field, and hence A_n is maximal.

Theorem 2.6. *Given a ring R, any ideal I is contained in some maximal ideal.*

Example 2.7. In l^∞, let $I_1 = \{(a_i) : a_i \neq 0 \text{ for only finitely many } i\}$ and $I_2 = \{(a_i) : \lim_{i\to\infty} a_i = 0\}$. These are both ideals in l^∞, since the sequences in l^∞ are bounded. By the above theorem there exist maximal ideals $J_1 \supseteq I_1$ and $J_2 \supseteq I_2$. Consider l^∞/J_1. Since J_1 is maximal, this is a field. Moreover, the map $\phi : \mathbb{Q} \to S/J_1$ given by $r \mapsto [(r, r, r, \ldots)]_{J_1}$. If $r \neq 0$ then $\phi(r) \neq [0]_{J_1}$. We will not prove it, but it is true, that the map ϕ extends to an isomorphism witnessing $S/J_1 \cong \mathbb{R}$. So we have a map $L : S \to \mathbb{R}$. It is a ring morphism, so $L((a_i)(b_i)) = L((a_i))L((b_i))$ and $L((a_i) + (b_i)) = L((a_i)) + L((b_i))$, and if (a_i) is a sequence with only finitely many non-zero entries, then $L((a_i)) = 0$, and from the construction of ϕ we see that if (a_i) is a constant sequence, say $a_i = r$ for all i, then $L((a_i)) = r$. Note that L has all of the good properties of normal limits, and generalizes normal limits. This is too good to be true. So where was the mistake? There is no mistake, but Theorem 2.6 depended on Zorn's Lemma (equivalent in ZF to the Axiom of Choice), and so we do not have a "hands-on" way of understanding L. One could redo the construction with J_2 as well, and get a similar generalization of lim.

Proof of Theorem 2.6. Recall that Zorn's Lemma asserts that if P is a partially ordered set in which every chain has an upper bound, then there exists a maximal element of P.

Consider $P = \{J \subsetneq R : I \subseteq J,\ J \text{ an ideal}\}$. P is partially ordered by inclusion. It is easy to check that if we have a chain in P then their union is a bound in P. So by Zorn's Lemma there is a maximal element of P, which is exactly a maximal ideal containing I. □

Definition 2.8. Let R be a ring. An ideal $P \subseteq R$ is called a **prime ideal** if for all $a, b \in R$, if $ab \in P$ then $a \in P$ or $b \in P$.

Example 2.9. In $\mathbb{Z}[x]$, then ideal $\langle x\rangle = \{f \in \mathbb{Z}[x] : f(0) = 0\}$ is prime. Indeed, if $fg \in \langle x\rangle$, so $f(x)g(x) = (fg)(x) = 0$, then either $f(x) = 0$ or $g(x) = 0$.

Theorem 2.10. *Let R be a ring, $P \subseteq R$ an ideal. Then P is a prime ideal if and only if R/P is a domain.*

Proof. Suppose that R/P is a domain. Suppose that $ab \in P$. Then $[a][b] = [ab] = [0]$ in R/P. Since R/P is a domain then $[a] = 0$ or $[b] = 0$. But that means either $a \in P$ or $b \in P$, so P is prime.

Conversely, suppose that P is prime. Suppose that $[a][b] = [0]$ in R/P. Then $[ab] = [0]$, so $ab \in P$. Since P is prime, $a \in P$ or $b \in P$. But this means that either $[a] = [0]$ or $[b] = [0]$. □

Theorem 2.11. *Let R be a ring, $M \subseteq R$ a maximal ideal. Then M is a prime ideal.*

Proof. Since M is maximal R/M is a field, hence a domain, so M is prime. □

Definition 2.12. Let R be a ring. An element $u \in R$ is called a **unit** if u is invertible. That is, if there exists $v \in R$ such that $uv = 1$. The collection R^* of all units in R forms a multiplicative group. If $a, b \in R$ are such that $a = ub$ for some unit u, then we say that a and b are **associates**.

Definition 2.13. Let R be a ring, $a, b \in R$. We say a **divides** b, and write $a \mid b$, if there exists $c \in R$ such that $ac = b$.

Proposition 2.14. *Let R be a domain, $a, b \in R$. If $a \mid b$ and $b \mid a$ then $a = ub$ where u is a unit in R.*

Proof. From the hypotheses we can write $b = ac$ and $a = bd$ for some $c, d \in R$. Thus $a = bd = acd$. So $a(1 - cd)0$. If $a = 0$ then $b = 0c = 0$, and we're done. If $a \neq 0$ we have $cd = 1$, so c and d are units. In particular, $a = bd$ and d is a unit. □

Definition 2.15. Let R be a ring. An element $p \in R$ such that $p \neq 0$ and p is not a unit (i.e., p is not invertible) is called a **prime** if $\langle p \rangle$ is a prime ideal. Equivalently, p is prime if whenever $p \mid ab$ then $p \mid a$ or $p \mid b$.

Definition 2.16. Let R be a ring. An element $p \in R$ such that $p \neq 0$ is called **irreducible** if whenever $p = ab$ then either a or b is a unit.

Note that the above two definitions generalize the notion of prime numbers (in $\mathbb{Z}$) in different ways. In particular, in $\mathbb{Z}$ the two definitions agree, provided that we allow for negative prime numbers. In general, the definitions genuinely are different, as we will see.

Proposition 2.17. *Let R be a domain, p a prime element. Then p is irreducible.*

Proof. Suppose that $p = ab$. Then $p \mid ab$, so without loss of generality $p \mid a$. So $a = pc$ for some $c \in R$. Then $p = pcb$. Since $p \neq 0$ we get $cb = 1$, so b is a unit. Thus p is irreducible. □

The converse to the above proposition is false:

Example 2.18. Let $R = \mathbb{Z}[\sqrt{-5}]$, the smallest subring of $\mathbb{C}$ containing $\mathbb{Z}$ and $\sqrt{-5}$. It is easy to check that $R = \left\{a + b\sqrt{-5} : a, b \in \mathbb{Z}\right\}$. Note that for any $a+b\sqrt{-5} \in R$, we have $\left|a + b\sqrt{-5}\right|^2 = a^2 + 5b^2 \in \mathbb{Z}$. The element $2 \in R$ is irreducible. To see this, note that $|2|^2 = 4$. Suppose that $2 = cd$ for some $c, d \in R$. Then $4 = |c|^2 |d|^2$. Both $|c|^2, |d|^2 \in \mathbb{Z}$, so $\|c\|^2$ is $1, 2$, or 4. Writing $c = a+b\sqrt{-5}$, we see from $|c|^2 = a^2 + 5b^2$ that $b = 0$. So $|c|^2 = a^2$. As $a \in \mathbb{Z}$ we get $|c|^2 \neq 2$. So either $|c|^2 = 1$ (in which case $c = \pm 1$) or $|c|^2 = 4$, in which case $4 = |c|^2 |d|^2$ implies $|d|^2 = 1$, so $d = \pm 1$ by the same argument as above. So 2 is irreducible.

On the other hand, $2 \mid 6$, and $6 = (1 + \sqrt{-5})(1 - \sqrt{-5})$. But $2 \nmid 1 + \sqrt{-5}, 2 \nmid 1 - \sqrt{-5}$, since $\frac{1+\sqrt{-5}}{2}, \frac{1-\sqrt{-5}}{2} \notin R$. So 2 is not prime in R.

Definition 2.19. Let R be a domain, $a, b \in R$. We say that q is **a greatest common divisor** of a, b if $q \mid a$, $q \mid b$, and if $q' \mid a$ and $q' \mid b$ then $q' \mid q$.

Note that even in $\mathbb{Z}$ greatest common divisors (in the above sense) are not unique. For example, -2 is a greatest common divisor of 6 and 10. However, we do have the following:

Proposition 2.20. *Let R be a domain, $a, b \in R$. If q, q' are both greatest common divisors of a and b then q and q' are associates.*

Proof. Since q, q' both divide a, b and are both greatest common divisors we get $q' \mid q$ and $q \mid q'$. So q, q' are associates by Proposition 2.14. □

In light of the above proposition, we write $q = \gcd(a, b)$ to mean that q is a greatest common divisor of a and b. This is well-defined up to associates. To say more about primes, we need to move to a more restricted class of rings.

Definition 2.21. A domain R is called a **unique factorization domain** if for any $a \in R$ such that $a \neq 0$ there exists a unit u and primes $p_i, \ldots, p_n$ such that $a = up_1 \cdots p_n$.

Proposition 2.22. *Let R be a unique factorization domain. Then factorizations are unique, in the sense that if $a = up_1 \cdots p_n = vq_1 \cdots q_m$ where the p_i's and q_j's are primes, and $u, v \in R^*$, then $n = m$ and there is some $\sigma \in S_n$ such that each q_i is an associate of $p_{\sigma(i)}$.*

Proof. Since $p_1 \cdots p_n = q_1 \cdots q_m$ then $p_1 \mid q_1 \cdots q_m$. So since p_1 is prime, there is some i such that $p_1 \mid q_i$. Permute the q's if necessary so that $i = 1$. Since q_1 is prime it is irreducible, so $p_1 \mid q_1$ implies that p_1, q_1 are associates. Thus $p_2 \cdots p_n = q_2 \cdots q_m u$ for some unit u. Now repeat. At the end the worst that could happen is we get $1 = q_{n+1} \cdots q_m$. But in this case q_{n+1} is a unit, so not a prime, contradiction. □

Proposition 2.23. *Let R be a unique factorization domain. Then $x \in R$ is irreducible if and only if x is prime.*

Proof. If x is prime then it is irreducible, by Proposition 2.17. Now suppose that x is irreducible. Since R is a unique factorization domain we have $x = up_1 \cdots p_n$ for some $u \in R^*$ and p_i primes. We this of this as a product $x = (p_1)(up_2 \cdots p_n)$. Since x is irreducible, one of the factors is a unit. Since p_1 is prime it is not a unit. But if $up_2 \cdots p_n$ is a unit then ? □

Theorem 2.24. *Let R be a ring. Then R is a unique factorization domain if and only if every non-zero element has a unique decomposition into irreducibles.*

Proof. It suffices to show that irreducible implies prime in a ring with unique decomposition into irreducibles. Suppose that $x \in R$ is irreducible. Suppose that $x \mid ab$. Write $a = a_1 \cdots a_n, b = b_1 \cdots b_m$ with the a_i, b_j's irreducible. So there exists z such that $xz = a_1 \cdots a_n \cdot b_1 \cdots b_m$. By the uniqueness of decompositions into irreducibles, x is one of the a_i's or one of the b_j's. If $x = a_i$ then $x \mid a$. If $x = b_i$ then $x \mid b$. □

Theorem 2.25. *Let R be a unique factorization domain, $a, b \in R$ then there exists a* gcd *for a and b.*

Proof. Write $a = up_1^{s_1} \cdots p_n^{s_n}$, $b = vp_1^{t_1} \cdots p_n^{t_n}$. Note that since only finitely many primes appear in the decomposition of each of a, b, and we allow $t_i, s_i = 0$, we may assume it is the same list of primes. A gcd for a, b is $p_1^{\min\{s_1,t_1\}} p_2^{\min\{s_2,t_2\}} \cdots p_n^{\min\{s_n,t_n\}}$. That this is a gcd is easy to check. □

So far we have not proved that any particular rings are unique factorization domains. We will show that $\mathbb{Z}$ is a unique factorization domain, as is $F[x]$ for any field F. We will do this by showing that if a ring has a norm then it is a principal ideal domain, and every principal ideal domain is a unique factorization domain.

Definition 2.26. A **Euclidean domain** is a domain R, along with a **norm**, which is a map $e : R \setminus \{0\} \to \mathbb{N}$ such that $e(ab) \geq e(a)$ and for every $a \in R$, $b \in R \setminus \{0\}$ there exist $q, r \in R$ such that $a = bq + r$ and either $r = 0$ or $e(r) < e(b)$.

Example 2.27.

1. $\mathbb{Z}$ is a Euclidean domain. The norm is $e(x) = |x|$.
2. $F[x]$, where F is a field, is a Euclidean domain. The norm is $e(f) = \deg(f)$. For example, if $a = x^3 - 2x^2 - 5x + 12$ and $b = x^2 + 1$, then we use polynomial long division to obtain $a = (x-2)b + (-6x+14)$, so $q = x - 2$, $r = -6x + 14$. Notice that $e(-6x+14) = 1 < 2 = e(b)$. Incidentally, $a(i) = -6i + 14 = r(i)$.

Definition 2.28. Let R be a ring. An ideal $I \subseteq R$ is called **principal** if there exists $x \in R$ such that $I = \langle x \rangle$. R is called a **principal ideal domain** if it is a domain in which every ideal is principal.

We will show later that if R is a principal ideal domain then it is a unique factorization domain. For the moment, let us show the following:

Theorem 2.29. *A Euclidean domain is a principal ideal domain.*

Proof. Let R be a Euclidean domain with norm e, $I \subseteq R$ an ideal. Let $x \in I$ be a non-zero element with lowest norm amongst elements of I. Then $I = \langle x \rangle$. Indeed, if $a \in I$, then since R is a Euclidean domain we can write $a = qx + r$ for some $q, r \in R$ with $r = 0$ or $e(r) < e(x)$. Write $r = a - qx$. Since $a, x \in I$ we get $r \in I$. By minimality of $e(x)$, we have $e(r) \not< e(x)$. Thus $r = 0$ so $a = qx \in \langle x \rangle$. Thus $I \subseteq \langle x \rangle$. Since $x \in I$, $\langle x \rangle \subseteq I$, so $I = \langle x \rangle$ as required. □

Proposition 2.30. *Let R be a principal ideal domain, $I \subseteq R$ a prime ideal. Then I is a maximal ideal.*

Proof. Since R is a principal ideal domain, we have $I = \langle p \rangle$ for some $p \in R$. Since I is prime this implies that p is a prime element. Suppose for a contradiction that J is an ideal such that $I \subsetneq J \subsetneq R$. Then $J = \langle x \rangle$ for some $x \in R$. So $p \in \langle x \rangle$. Thus $p = ax$ for some $a \in R$. Since p is prime, either $p \mid a$ or $p \mid x$. If $p \mid a$ then $a = bp$ for some $b \in R$. Then $p = bpx$, and since R is a domain this implies $1 = bx$. Thus x is a unit, so $J = \langle x \rangle = R$, contradicting our choice of J. On the other hand, if $p \mid x$ then since already we have $x \mid p$, so we get $p = ux$ for some unit u by Proposition 2.14. Then $I = \langle p \rangle = \langle x \rangle = J$, again contradicting our choice of J. So in both cases we got a contradiction, and we're done. □

Definition 2.31. Let R be a ring. We say that R is **Noetherian** if any increasing sequence $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ of ideals is eventually constant.

Proposition 2.32. *Let R be a principal ideal domain. Then R is Noetherian.*

Proof. Let $I_1 \subseteq I_2 \subseteq \ldots$ be an increasing sequence of ideals in R. Then $\bigcup_{k<\omega} I_k$ is again an ideal of R. Thus $\bigcup_{k<\omega} I_k = \langle x \rangle$ for some $x \in R$, since R is a principal ideal domain. So $x \in I_n$ for some n. But then $\langle x \rangle \subseteq I_n$, so $\bigcup_{k<\omega} I_k = I_n$. □

Theorem 2.33. *Let R be a principal ideal domain. Then R is a unique factorization domain.*

Proof. Pick $x_1 \in R$, $x_1 \neq 0$. Find a maximal ideal $M_1 \subset R$ such that $\langle x_1 \rangle \subseteq M_1$. Since R is a principal ideal domain, $M_1 = \langle p_1 \rangle$ for some prime $p_1 \in R$. So $x_1 = p_1 x_2$ for some $x_2 \in R$, since $x \in M_1$. Find a maximal M_2 such that $\langle x_2 \rangle \subseteq M_2$. Again $M_2 = \langle p_2 \rangle$ for a prime p_2. Then $x_2 = p_2 x_3$ for some $x_3 \in R$. Repeat. Notice that this process stops when some x_{n+1} is a unit. In this case, we get $x_1 = p_1 x_2 = p_1 p_2 x_3 = \ldots = p_1 \cdots p_n x_{n+1}$, and the last expression is the required factorization of x_1. If the process never stops then we get a sequence $\langle x_1 \rangle \subseteq \langle x_2 \rangle \subseteq \langle x_3 \rangle \subseteq \ldots$. Since R is Noetherian by the previous Proposition, we get that there exists n such that $\langle x_i \rangle = \langle x_n \rangle$ for all $i \geq n$. In particular, x_n and x_{n+1} are associates. But $x_n = p_n x_{n+1}$, so p_n is a unit. But we said that p_n is a prime, contradiction. So in fact the process stops, and we're done. □

Theorem 2.34. *In a principal ideal domain, $\langle a, b \rangle = \langle \gcd(a, b) \rangle$. (More precisely, $\langle a, b \rangle = \langle q \rangle \iff q$ is a* gcd *of a, b). Thus there exist $s, t \in R$ such that* $\gcd(a, b) = sa + tb$.

Proof. It suffices to show that $\langle a, b \rangle = \langle q \rangle \iff q$ is a gcd of a, b. First, suppose that $\langle a, b \rangle = \langle q \rangle$. Then $a \in \langle q \rangle$, so $q \mid a$. Similarly, $q \mid b$. Suppose that $q' \mid a$ and $q' \mid b$. Then since $\langle q \rangle = \langle a, b \rangle$ we have $q = sa + tb$ for some $s, t \in R$. Since $q' \mid a$ and $q' \mid b$, we have $q' \mid q$. So we have shown q is a gcd of a and b.

We omit the converse. □

Example 2.35. Consider $F[x, y]$, the polynomial ring in two variables over a field F. This ring is a unique factorization domain (though we will not prove it), but it is not a principal ideal domain. Indeed, $\langle x, y \rangle$ is not a principal ideal in this ring. This is because $\gcd(x, y) = 1$, and there are no $s, t \in F[x, y]$ such that $sx + ty = 1$.

Recall that in $\mathbb{Z}$ we have an effective way, given a, b, of finding s, t such that $sa + tb = \gcd(a, b)$, namely the Euclidean algorithm. This extends to arbitrary Euclidean domains. Without loss of generality, $e(a) \geq e(b)$. If $b \mid a$ then $\langle a, b \rangle = \langle b \rangle$, so $b = \gcd(a, b)$. We may then take $s = 0, t = 1$. Otherwise, write $a = bq + r$, where $e(r) < e(b)$. Then $\langle a, b \rangle = \langle b, r \rangle$. Indeed, we have $a = bq + r \in \langle b, r \rangle$, and $r = a - bq \in \langle a, b \rangle$. So now it suffices to solve the problem for b, r. By recursion, find s', t' such that $\gcd(a, b) = \gcd(b, r) = s'b + t'r$. Then $\gcd(a, b) = s'b + t'(a - bq) = t'(s' - q)b + t'a$. The procedure is guaranteed to stop, since at each stage we have $e(r) < e(b)$.

Given a principal ideal domain R, and a non-zero $x \in R$, define $d(x)$ to be x with every prime replaced by 2. That is, write $x = p_1 \cdots p_n$ (which can be done with n unique, since R is a unique factorization domain). Then set $d(x) = 2^n$. Then $d : R \backslash \{0\} \to \mathbb{N} \backslash \{0\}$. This can be extended by $d(0) = 0$. We have $d(xy) = d(x)d(y)$ if both are non-zero. In particular, $d(xy) \geq d(x)$. If $a, b \neq 0$ then either $a \in \langle b \rangle$, in which case we have nothing to say, or $\langle q \rangle = \langle a, b \rangle \supsetneq \langle b \rangle$, where $q = \gcd(a, b)$. Then $d(q) < d(b)$. There exist s, t such that $q = sa + tb$. So $d(sa + tb) < d(b)$. In this situation $sa + tb$ plays the role of r in a Euclidean domain. This is the only way in which d fails to be a Euclidean norm.

Definition 2.36. A **Dedekind-Hasse norm** is a function $d : R \to \mathbb{N}$ such that for any $a, b \neq 0$, we have either $a \in \langle b \rangle$, or there exist $q \in \langle a, b \rangle$ such that $d(q) < d(b)$.

Theorem 2.37. *A ring R is a principal ideal domain if and only if R has a Dedekind-Hasse norm.*

Proof. In the above paragraph we showed that a principal ideal domain has a Dedekind-Hasse norm, and we even constructed it. If R has a Dedekind-Hasse norm, we use the same proof as Theorem 2.33, mutatis mutandis. □

3. Modules

Definition 3.1. Let R be a ring. A **module** over R is a set with operations satisfying exactly the same axioms as the axioms for a vector space over a field. More precisely, a module is an abelian group M along with a map $\times : R \times M \to M$ such that for any $x, y \in R, m, n \in M$:

(1) $(x + y)m = xm + ym$
(2) $x(m + n) = xm + xn$
(3) $0m = 0$
(4) $x(ym) = (xy)m$
(5) $1m = m$

Example 3.2. (1) Let F be a field. A module over F is exactly a vector space over F.

(2) A module over $\mathbb{Z}$ is exactly the same thing as an abelian group. The scalar multiplication is $nx = x + x + \ldots + x$ (n copies) if $n > 0$, and $nx = -x - x \ldots - x$ (n copies) if $n < 0$.
(3) Given a vector space V over a field F and a linear map $T : V \to V$, we can make V into a module over $F[x]$ by $(\sum a_i x^i)v = \sum a_i T^i(v)$. It is easy to check that this action satisfies the definition of a module.
(4) Given an ideal $I \subseteq R$, R/I is an R-module in the obvious way, by $r[r'] = [rr']$.
(5) If R is non-commutative there are two notions of modules, namely left and right R-modules. R-mod is the category of left R-modules (so the ring action is on the left), and mod-R is the category of right R-modules, where the action of R is on the right. So our above definition is technically the definition of a left R-module, but the corresponding definition of a right R-module is clear. We get $m(ab) = (ma)b$ as an axiom, which is not the same as the corresponding axiom $(ab)b = a(bm)$. So in general, left R-modules are not the same as right R-modules. Unless otherwise stated, we assume modules are left modules.
(6) Consider column vectors of length n with coefficients in a (non-commutative) ring R. This is R^n. This is a module over R (on the left or right). It is also a module over $M_{n\times n}(R)$, using matrix multiplication. This is a left R-module with the obvious action, the matrix goes on the left. On the other hand, the row vectors are the elements of $(R^n)^T$. This is a right $M_{n\times n}(R)$-module, again using matrix multiplication.

Definition 3.3. A **morphism of R-modules** is a function $f : M \to N$ (M, N both R-modules) such that f is a homomorphism of the underlying abelian groups, and respects the action of the ring, in the sense that $f(rm) = rf(m)$ for all $r \in R, m \in M$. We think of these as analogous to linear transformations of vector spaces.

With the above morphisms, R-mod forms a category.

Definition 3.4. The a **submodule** is a subset closed under all of the relevant operations.

Definition 3.5. Given a morphism $f : M \to N$ we have $\ker f = \{m \in M : f(m) = 0\}$, a submodule of M. We also have $\operatorname{im} f = \{n \in N : \exists m \in M f(m) = n\}$ is a submodule of N.

Given $N \subseteq M$, we construct M/N in the obvious way.

Theorem 3.6 (Isomorphism Theorems)**.** *(1) If $\phi : M \to N$ is a morphism then $M/\ker\phi \cong \operatorname{im}\phi$.*

(2) If $A, B \subseteq M$ then $A + B/B \cong A/A \cap B$.

(3) If $A \subseteq B \subseteq M$ then $(M/A)/(B/A) \cong M/B$.

(4) The fourth (lattice) isomorphism theorem also holds.

Definition 3.7. Given R-modules M, N, we construct another R-module $M \bigoplus N = \{(m, n) : m \in M, n \in N\}$ with pointwise operations inherited from M and N.

Theorem 3.8. *Let M, N be R-modules. Let $i_M : M \to M \bigoplus N$ be the morphism $m \mapsto (m, 0)$ and $i_N : N \to M \bigoplus N$ be $n \mapsto (0, n)$. Given morphisms $\phi : M \to P$ and $\psi : N \to P$ there is a unique $\lambda : M \bigoplus N \to P$ such that $\lambda \circ i_M = \phi$ and $\lambda \circ i_N = \psi$. The map is $\lambda(m, n) = \phi(m) + \psi(n)$.*

Moreover, if Z is any other R-module with the above property, then $Z \cong M \bigoplus N$. Thus $M \bigoplus N$ is the categorical direct sum of M and N.

Theorem 3.9. *Let M, N be R-modules. Then there are projection morphisms $\pi_M : M \bigoplus N \to M$ and $\pi_N : M \bigoplus N \to N$ given in the obvious ways. If P is an R-module such with morphisms $\phi : P \to M$ and $\psi : P \to N$ then there is a unique $\lambda : P \to M \bigoplus N$ such that $\pi_M \circ \lambda = \phi$ and $\pi_N \circ \lambda = \psi$. If Z is any other R-module with the above property, then $Z \cong M \bigoplus N$. Thus $M \bigoplus N$ is the categorical direct product of M and N.*

Both the notion of categorical direct product and categorical direct sum extend to infinitely many factors, but there they do not coincide. We will not need this result.

Example 3.10. (1) If V, W are vector spaces then we know that $\dim(V \bigoplus W) = \dim(V) + \dim(W)$.

(2) If $a, b \in R$ with R a domain, and $\gcd(a, b) = 1$, and there exist $s, t \in R$ such that $sa + tb = 1$ (for example, we saw that this happens if R is a PID), then $R/\langle a\rangle \bigoplus R/\langle b\rangle \cong R/\langle ab\rangle$. This is the Chinese Remainder Theorem. In fact the statement is true as rings, but we will for the moment only prove it as modules.

Proof. We need to construct a map $\phi : R/\langle a\rangle \bigoplus R/\langle b\rangle \to R/\langle ab\rangle$ and a map $\psi : R/\langle ab\rangle \to R/\langle a\rangle \bigoplus R/\langle b\rangle$, and show that the maps are morphisms and inverses to each other. It suffices to give maps on (resp. to) each factor of the direct sum, by what we saw above. So we define $\phi_1 : R/\langle a\rangle \to R/\langle ab\rangle$ by $\phi_1([r]) = [tbr]$ and $\phi_2 : R/\langle b\rangle \to R/\langle ab\rangle$ by $\phi_2([r']) = [r'sa]$. An above theorem gives the desired map ϕ, and it is easy to check that it is a morphism. On the other hand, we use multiplication by 1 maps for ψ_1, ψ_2. Then it is again easy to check that ψ given by the above theorem is a morphism, and that ψ, ϕ are inverses. The details are an exercise. □

Now we define a product of modules:

Definition 3.11. Let M, N be R-modules. Define $M \bigotimes_R N = \left\{ \sum_{i=1}^{k} a_i m_i \otimes n_i : a_i \in R, m_i \in M, n_i \in N \right\} /\mathrm{rela}$
Here $m \otimes n$ is just a symbol, which we think of as representing (m, n), and the sum is a formal sum. The relations are as follows. Consider the set-theoretic Cartesian product $M \times N$, and the map $\phi : M \times N \to M \bigotimes N$ given by $(m, n) \mapsto m \otimes n$. The relations are exactly the ones making ϕ bilinear. So we need $(am) \otimes n = a(m \otimes n) = m \otimes (an)$. This is the first of the relations. The others are that $(m_1 + m_2) \otimes n = m_1 \otimes n + m_2 \otimes n$ and $m \otimes (n_1 + n_2) = m \otimes n_1 + m \otimes n_2$.

Example 3.12. Let $R = F$ be a field, $M = V, N = W$ be vector spaces over F. Let $\{v_i\}$ be a basis for V, and $\{w_j\}$ be a basis for W. Then $\{v_i \otimes w_j\}$ is a basis for $V \bigotimes_F W$. In particular, if V, W are finite-dimensional then $\dim(V \bigotimes_F W) = \dim(V) \dim(W)$.

Proof. We need to prove that $\{v_i \otimes w_j\}$ is a basis for $V \bigotimes_F W$. First, consider any $\sum a_\alpha f_\alpha \otimes g_\alpha$ where $f_\alpha \in V, g_\alpha \in W$. Write $f_\alpha = \sum b_{\alpha,i} v_i$. Similarly, $g_\alpha = \sum c_{\alpha,j} w_j$. So

$$\begin{aligned} \sum a_\alpha f_\alpha \otimes g_\alpha &= \sum a_\alpha \left(\sum b_{\alpha,i} v_i \right) \otimes \left(\sum c_{\alpha,j} w_j \right) \\ &= \sum_\alpha \sum_i \sum_j a_\alpha b_{\alpha,i} c_{\alpha,j} (v_i \otimes w_j) \end{aligned}$$

Thus $\{v_i \otimes w_j\}$ spans $V \bigotimes_F W$.

Now we need to do linear independence. Let $\{\phi_i\}, \{\psi_j\}$ be the dual bases of $\{v_i\}$ and $\{w_j\}$ in V^* and W^*, respectively.

Claim 3.13. If $\phi \in V^*$ and $\psi \in W^*$ then $\phi \otimes \psi : V \bigotimes_F W \to F$ given by $\phi \otimes \psi(\sum a_\alpha v_\alpha \otimes w_\alpha) = \sum a_\alpha \phi(v_\alpha) \psi(w_\alpha)$ is well-defined.

The above claim is easy to verify, and just involves checking that the relations quotiented out by in constructing $V \bigotimes_F W$ are preserved. It is clear that $\phi \otimes \psi$ is linear.

Now assume that $\sum a_{i,j} v_i \otimes w_j = 0$. Apply $\phi_{i'} \otimes \psi_{j'}$ to both sides. We get $\sum a_{i,j} \delta_{i,i'} \delta_{j,j'} = 0$, so $a_{i',j'} = 0$ and we got linear independence. □

Claim 3.14. Let R be a ring. If $q = \gcd(a, b)$ and $q = sa + tb$ then $R/\langle a \rangle \bigotimes_R R/\langle b \rangle \cong R/\langle q \rangle$.

Proof. Define $\phi : R/\langle a \rangle \bigotimes_R R/\langle b \rangle \to R/\langle q \rangle$ by $[r_1]_a \otimes [r_2]_b \mapsto [r_1 r_2]_q$, extended linearly. We need to check that this is well-defined, but this is easy. Define also $\psi : R/\langle q \rangle \to R/\langle a \rangle \bigotimes_R R/\langle b \rangle$ by $[r]_q \mapsto [r]_a \otimes [1]_b = [1]_a \otimes [r]_b$. Again we need to check that this is well-defined. It suffices to note that $[0]_q = [q]_q = [sa + tb] = [sa] + [tb] \mapsto [sa]_a \otimes [1]_b + [1]_a \otimes [tb]_b = 0$. Then it is easy to check that the composition is the identity. □

Example 3.15. Let F be a field, and consider $F[x]$. It is a module over F. Similarly, $F[y]$ is an F-module. By a previous exercise we can compute a basis, and hence find that $F[x] \bigotimes_F F[y] \cong F[x, y]$.

Example 3.16. This example is not quite right, but is morally true. Suppose that X, Y are topological spaces. Then $C(X)$, the continuous functions from X to $\mathbb{R}$ is a real vector space. Similarly for $C(Y)$. So we can ask what is $C(X) \bigotimes_{\mathbb{R}} C(Y)$? Polynomials are dense in the space of continuous functions, so the previous example says that, morally speaking, we might expect that $C(X) \bigotimes_{\mathbb{R}} C(Y)$“ $\cong$ ”$C(X \times Y)$.

But this argument used analytic techniques (density), which we cannot apply in general. There is, however, an injective map $C(X) \bigotimes_{\mathbb{R}} C(Y) \to C(X \times Y)$, and if X and Y are sufficiently well-behaved (for example, compact Hausdorff suffices, lesser conditions might also do) then the image of this map is dense.

Proposition 3.17. *There is a bilinear map $\iota : M \times N \to M \bigotimes N$, given by $\iota(m, n) = m \otimes n$. If P is another R-module and there is a bilinear map $F : M \times N \to P$ then there exists a unique module homomorphism $\tilde{f} : M \bigotimes_R N \to P$ such that $\tilde{f} \circ \iota = F$.*

Proof. We must define $\tilde{f}(m \otimes n) = F(m, n)$. Then we must define $\tilde{f}(\sum_{i=1}^{n} a_i m_i \otimes n_i) = \sum_{i=1}^{n} a_i F(m_i, n_i)$. Since we had no choice this establishes uniqueness and the relation $\tilde{f} \circ \iota = F$. We only need to know that $\tilde{f}$ is well-defined. But this is clear because F is bilinear. □

Proposition 3.18. *The property in Proposition 3.17 determines $M \bigotimes_R N$. That is, if we had another module $M \widehat{\bigotimes} N$ is such that there is a bilinear map $\iota' : M \times N \to M \widehat{\bigotimes} N$ such that if $f : M \times N$ is bilinear then there exists a unique $\tilde{f} : M \widehat{\bigotimes} N \to P$ making the diagram commute, then $M \widehat{\bigotimes} N \cong M \bigotimes_R N$.*

Proof. In fact there is a unique isomorphism respecting the ι's, but we will not show that here.

Let $P = M \bigotimes_R N$. Then there is a map $\iota : M \times N \to P$, namely ι. There is also $\iota' : M \times N \to M \widehat{\bigotimes} N$. By the two universal properties we get unique homomorphisms $M \bigotimes_R N \to M \widehat{\bigotimes} N$ and $M \widehat{\bigotimes} N \to M \bigotimes_R N$, and it is an easy exercise to see that these maps are inverse to one another. □

Theorem 3.19. *$(R - mod, \bigoplus, \bigotimes, 0, R)$ is a "ring", in as much as this makes sense. More precisely,*

(1) $(M \bigoplus N) \bigoplus P \cong M \bigoplus (N \bigoplus P)$.
(2) $(M \bigotimes_R N) \bigotimes_R P \cong M \bigotimes_R (N \bigotimes P)$.
(3) $(M \bigoplus N) \bigotimes_R P \cong M \bigotimes_R P \bigoplus N \bigotimes_R P$.
(4) $M \bigotimes_R N \cong N \bigotimes_R M$.
(5) $M \bigoplus 0 \cong M$.
(6) $M \bigotimes_R 0 \cong 0$.
(7) $M \bigotimes_R R \cong M$.

Proof. We just show the last claim. Define a map by $m \otimes r \mapsto rm$, extended linearly. Define also $m \mapsto m \otimes 1$. It is easy to check that both maps are well-defined and inverse to each other. □

Example 3.20. $\mathbb{Q}$ is a $\mathbb{Z}$-module in an obvious way, as is $\mathbb{Z}^n$. Thus $\mathbb{Z}^n \bigotimes_{\mathbb{Z}} \mathbb{Q}$ is also a $\mathbb{Z}$-module. We can calculate which one it is:

$$\begin{aligned} \mathbb{Q} \bigotimes_{\mathbb{Z}} \mathbb{Z}^n &\cong (\mathbb{Q} \bigotimes_{\mathbb{Z}} \mathbb{Z})^n \\ &\cong \mathbb{Q}^n \end{aligned}$$

More generally, if $\phi : R \to S$ is a morphism of rings, then $M_S := S \bigotimes_R M$ is an S-module. Indeed, the action is $s(s' \otimes m) = (ss' \otimes m)$. We say that M_S is obtained from M by **extension of scalars**. In particular, $(R^n)_S = S^n$, by the same calculation as we did above.

Claim 3.21. Over the ring $R = \mathbb{Z}$, if $M \cong \mathbb{Z}^k \oplus \bigoplus \mathbb{Z}/\langle n_i \rangle$, then k is unique.

Proof. Consider $M_{\mathbb{Q}}$, a $\mathbb{Q}$-module, that is, a $\mathbb{Q}$-vector space. So it has some well-defined dimension. Moreover, for any n, we have $\mathbb{Q} \bigotimes_{\mathbb{Z}} \mathbb{Z}/\langle n \rangle = 0$. Indeed, a basic tensor is $a \otimes [j]_n = n\frac{a}{n} \otimes [j]_n = \frac{a}{n} \otimes [nj]_n = \frac{a}{n} \otimes [0] = 0$ Thus:

$$\begin{aligned}\dim(M_{\mathbb{Q}}) &= \dim\left(\mathbb{Q}^k \oplus \bigoplus\left(\mathbb{Q} \bigotimes_{\mathbb{Z}} \mathbb{Z}/\langle n_i \rangle\right)\right)\\ &= \dim(\mathbb{Q}^k)\\ &= k\end{aligned}$$

□

Definition 3.22. A **functor** is a map $F : \mathcal{C} \to \mathcal{D}$ where $\mathcal{C}, \mathcal{D}$ are categories, such that if $\phi : A \to B$ is a morphism then there is a morphism $F\phi : FA \to FB$, in such a way that $F(\phi \circ \psi) = F\phi \circ F\psi$. Moreover, the identity morphisms are mapped to identity morphisms.

Example 3.23. Let $\mathcal{C}$ be the category of groups with their homomorphisms, and let $\mathcal{S}$ be the category of sets with their functions, and let $F : \mathcal{C} \to \mathcal{S}$ be the functor which "forgets" the group structure. One can construct many similar examples of **forgetful functors**.

Let $\mathcal{C}$ be the category of pointed topological spaces with their basepoint-preserving continuous maps. Let $\mathcal{D}$ be the category of groups with homomorphisms. Then the functor F which takes (X, p) to its fundamental group is a functor.

Definition 3.24. A **bifunctor** is a map $F : \mathcal{C} \times \mathcal{D} \to \mathcal{E}$, where $\mathcal{C}, \mathcal{D}, \mathcal{E}$ are categories, such that F is a functor in each variable separately.

Example 3.25. $\bigotimes$ is a bifunctor. That is, fix a module N, then the map $M \mapsto M \bigotimes N$ is a functor, and similarly if we fix a module M then the map $N \mapsto M \bigotimes N$ is also a functor. In more detail, suppose that $M_1 \to^f M_2$. Then there is a map $f \otimes N : M_1 \bigotimes N \to M_2 \bigotimes N$, which is given by the linear extension of $m_1 \otimes n \mapsto f(m_1) \otimes n$. One needs to check that this is well-defined, but this is not difficult, since f is a module morphism. One also needs to check that if $M_1 \mapsto^g M_2 \mapsto^f M_3$, then $(f \circ g) \otimes N = f \otimes N \circ g \otimes N$. This is also obvious. Note that if we have morphisms $f : M_1 \to M_2$ and $g : N_1 \to N_2$ then there is a map $f \otimes g : M_1 \bigotimes N_1 \to M_2 \bigotimes N_2$, given by the linear extension of $m_1 \otimes n_1 \mapsto f(m_1) \otimes g(n_1)$.

Now we can return to our main goal, the structure theorem for finitely generated modules over Principal Ideal Domains.

Definition 3.26. An R-module M is **finitely generated** if there exist $x_1, \ldots, x_n \in M$ such that the map $R^n \to^\rho M$ given by $(r_1, \ldots, r_n) \mapsto \sum_{i=1}^n r_i x_i$ is surjective.

We now sketch the existence part of the theorem we are aiming for, and later we will do the details.

If we are lucky there will be some finite c such that the map $R^c \mapsto^A R^n$ is a surjection into the kernel of ρ. We will later see that in fact this always happens. We can think of A as an $n \times c$ matrix over R. A contains a complete description of M, since by the first isomorphism theorem $M \cong R^n / \operatorname{im}(A)$. A is not unique, since the same is true of any A' obtained from A by $A' = PAQ$ where P, Q are invertible. Over a field multiplying by P on the left is equivalent to performing row operations, and multiplying by Q on the right is equivalent to performing column operations. So the question is, given $A \in M_{n\times c}(R)$, how nice a form can A be given by performing row and column operations?

Let us think about the case when we are over a field, $A = (a_{i,j})$. If every $a_{i,j}$ is 0 then we are happy, and in fact then $M \cong R^n$. Otherwise, we can put the matrix into a form where the first several entries on the diagonal are 1, and the rest of the matrix is 0.

Now we do the same over a ring R. Again if everything is 0 then there is nothing to do. Unfortunately, if we have a row $(a \cdots b \cdots)$ we cannot add $-b/a$ times column 1 to get rid of the b. But suppose that we are in a Euclidean domain, and have a row $(6 \cdots 40 \cdots)$. We cannot get rid of the 40 directly, but we can subtract $6\times$ column 1 to get $(6 \cdots 4 \cdots)$, and swapping columns $(4 \cdots 6 \cdots)$. Again we can do this process to get $(4 \cdots 2 \cdots)$, and flipping, $(2 \cdots 4 \cdots)$. Now we can get rid of the 4, and get $(2 \cdots 0 \cdots)$. Repeating the same process we can eventually replace each row $(a_1 a_2 \cdots a_n)$ by $(q0 \cdots 0)$ where $q = \gcd(a_1, \ldots, a_n)$. Do the same over columns. We eventually come to a matrix with q it the top left corner, 0's elsewhere on the first row and column, and q divides every entry of the remainder of the matrix. Now we can repeat, never touching the first row and column again. Eventually we get to a matrix with a_i's in the first m places of the diagonal, and the rest 0, and such that $a_1 \mid a_2 \mid a_3 \cdots \mid a_m$. Call this matrix A'. We then have $M \cong R^n/\operatorname{im}(A') = \langle e_1, \ldots, e_n\rangle/\langle a_1e_1, a_2e_2, \ldots, a_me_m\rangle = R^{n-m} \bigoplus R/\langle a_1\rangle \bigoplus \cdots \bigoplus R/\langle a_m\rangle$. We already saw that if $\gcd(a, b) = 1$ then $R/\langle a\rangle \bigoplus R/\langle b\rangle \cong R/\langle ab\rangle$, so using this and factoring the a_i's into primes we can turn the above direct sum into the desired form. Now let us do this rigourously.

Theorem 3.27. *If M is a finitely generated module over a principal ideal domain R then M is isomorphic to a direct sum*

$$M \cong R^k \bigoplus \bigoplus R/\langle p_i^{s_i}\rangle$$

where each $p_i \in R$ is prime and $s_i \in \mathbb{R}$.

Proof. M can be written as $\langle x_1, \ldots, x_n\rangle/\langle a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n = 0, \ldots\rangle$. Let r_1 be the relation $a_{1,1}x_1 + \cdots + a_{1,n}x_n = 0$, and so on, so $\{r_i : i < \kappa\}$ is the set of all the relations being quotiented by, for some (possibly infinite) cardinal κ. Among all such presentations of M using precisely n generators consider one in which $a_{1,1} \neq 0$, and the Dedekind-Hasse norm $d(a_{1,1})$ is minimal. Without loss of generality, $r_1 = a_{1,1}x_1 + 0$ and $a_{i,1} = 0$ for all $i < \kappa$. This is the key step of the proof, and needs some justification. First,

Claim 3.28. $a_{1,1} \mid a_{1,j}$ for all j, and $a_{1,1} \mid a_{i,1}$ for all i.

Proof. Let $z = a_{1,1}$, and $b = a_{1,j}$ be such that $x \not| y$. Let $x = x_1, y = x_j$. Then $r_1 = ax + by + \cdots$. We are in a PID, so we have $q = \gcd(a, b) = ta + sb$ for some $t, s \in R$. Let $x' = \frac{a}{q}x + \frac{b}{q}y$ and $y' = -tx + sy$. Then $x = sx' - \frac{b}{q}y'$ and $y = tx' + \frac{a}{q}y'$. So we may write $r_1 = ax + by + \cdots = a(sx' - \frac{b}{q}y') + b(tx' + \frac{a}{q}y') + \cdots$, where the other terms do not involve x and y, hence not x', y' either. Expanding, we get $r_1 = qx' + 0y' + \cdots$. So we found a presentation of the module in which the first coefficient is less than the one we started with, contradicting our original choice of presentation. This proves the first assertion.

Now suppose that $r_1 = ax + \cdots$, and $r_2 = bx + \cdots$ (here r_2 is any other relation). We claim that $a \mid b$. If not, let $q = \gcd(a, b)$, which has smaller norm than a. Write

$q = sa + tb$. Then replace r_1, r_2 by $r_1' = sr_1 + tr_2, r_2' = -\frac{b}{q}r_1 + \frac{a}{q}r_2$. It is easy to check that r_1, r_2 are equivalent to r_1', r_2'. But here $r_1' = sax + btx + \cdots = qx + \cdots$, contradicting minimality of a again. □

Now we finish the proof of the key step. We have $a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n, \cdots, a_{\infty,1}x_1 + \cdots + a_{\infty,n}x_n$, where we write $a_{\infty,i}$ to emphasize that there may be infinitely many relations. By the claim $a_{1,1} \mid a_{1,i}$ for all i and $a_{1,1} \mid a_{j,1}$ for all j. By standard column operations we can replace these relations by $a_{1,1}x_1 + 0 + \cdots + 0$, $a_{2,1}x_1 + *, \ldots$, where the $*$s are non-zero. We still have $a_{1,1} \mid a_{j,1}$ for all j, so adding multiples of the first row to each other row (possibly infinitely many operations) we get to the form we claimed. That is, we get relations which can be expressed as the (infinite) matrix:

$$\begin{pmatrix} a_{1,1} & 0 & 0 \\ 0 & * & * \\ 0 & * & * \end{pmatrix}$$

Where here we do not know what the $*$s are. The operation is well-defined, and moreover it does not change the span of the rows. Moreover, $a_{1,1} \mid *$ for any $*$ in the rest of the matrix. Continuing by induction, after a finite number of steps we get the following form, with 0's elsewhere.

$$\begin{pmatrix} a_{1,1} & 0 & 0 & 0 \\ 0 & a_{2,2} & 0 & 0 \\ 0 & 0 & a_{m,m} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Thus $M = R^n / \langle a_{1,1}e_1 = 0, \cdots, a_{m,m}e_m = 0\rangle$, and we finish as before. □

We next will show uniqueness. The ideas are the same as in Claim 3.21. To start,

Proposition 3.29. *Let R be a domain. Then there is a unique (up to isomorphism) field $Q(R)$, called the* ***field of fractions of*** *R, such that there is an injection $\iota : R \to Q(R)$ such that if $\phi : R \to F$ is a ring morphism and F is a field then there exists a unique map $\psi : Q(R) \to F$ such that $\psi \circ \iota = \phi$. $Q(R)$ looks like $\left\{\frac{r}{s} : r, s \in R, s \neq 0\right\}$.*

We will come back to proving this, after a brief digression:

Definition 3.30. We say that $S \subseteq R \setminus \{0\}$ is **multiplicative** if for all $a, b \in S$ we have $ab \in S$, and $1 \in S$.

Proposition 3.29 will follow immediately from the following more general claim applied to $S = R \setminus \{0\}$.

Proposition 3.31. *Let R be a domain, and $S \subseteq R \setminus \{0\}$ be a multiplicative set. Then there exists a unique (up to isomorphism) ring $S^{-1}R$ and an injective ring morphism $\iota : R \to S^{-1}R$, such that any map $R \to A$ which sends $S \to A^*$ factors through $S^{-1}R$.*

Example 3.32.

1. $S = R \setminus \{0\}$. Then $S^{-1}R = Q(R)$ from Proposition 3.29.
2. Let $P \subseteq R$ be a prime ideal, and let $S = R \setminus P$. Then S is multiplicative. Indeed, if $a, b \in S$, then $a, b \notin P$. Then $ab \notin P$, since if $ab \in P$ then since P is prime we would get $a \in P$ or $b \in P$. In this case $S^{-1}P$ is called the **localization of** R **at** P.

(3) Let $R = \mathbb{Z}$, and let $S = \{2^n : n \in \mathbb{N}\}$. Then $S^{-1}R = \left\{\frac{a}{2^k} : a \in \mathbb{Z}, k \in \mathbb{N}\right\}$. This is the ring of **dyadic rationals**. One can, of course, use a number other than 2. Note that this is not the same thing as the 2-adic numbers.

Proof Sketch of Proposition 3.31. We define $S^{-1}R = \{(r,s) : r \in R, s \in S\} / \sim$, where $\sim$ is given by $(r_1, s_1) \sim (r_2, s_2) \iff r_1 s_2 = r_2 s_1$. It is clear that the relation $\sim$ is symmetric and reflexive, but transitivity needs to be checked. Suppose $(r_1, s_1) \sim (r_2, s_2) \sim (r_3, s_3)$. Then $r_1 s_2 = r_2 s_1$ and $r_2 s_3 = r_3 s_2$. Thus we can multiply to get $r_1 s_2 s_3 = r_2 s_1 s_3$ and $r_2 s_3 s_1 = r_3 s_2 s_1$. Thus $r_1 s_2 s_3 = r_3 s_2 s_1$, so $r_1 s_3 = r_3 s_1$ since we are in a domain. So $(r_1, s_1) \sim (r_3, s_3)$, as claimed. So $\sim$ really is an equivalence relation.

Now we need to define operations to make this set into a ring. We define the 0 element to be $(0, 1)$. 1 will be $(1, 1)$. We define $(a, b) + (c, d) = (ad + bc, bd)$, and $(a, b)(c, d) = (ac, bd)$. It is easy to check that all of these are well-defined. Then we need to check that this satisfies the axioms of a commutative ring, but this is also straightforward. The map $\iota : R \to S^{-1}R$ is given by $\iota(r) = [(r, 1)]$. We need to check that ι is injective. Suppose $\iota(r_1) = \iota(r_2)$. Then $[(r_1, 1)] = [(r_2, 1)]$, so $r_1 \cdot 1 = 1 \cdot r_2$, so $r_1 = r_2$ and ι is injective. It is just as easy to check that ι is a ring morphism. Finally, we need to see that maps $R \to A$ with $S \to A^*$ factor uniquely through $S^{-1}R$. But this is also clear, as are all of the uniqueness claims. □

Proposition 3.33. *Suppose that $M \cong R^k \oplus \bigoplus_{i=1}^n R/\langle p_i \rangle^{s_i}$. Then the right side is uniquely determined by M.*

Proof. First, observe that (as in Claim 3.21 we have:

$$\begin{aligned}\dim(M_{Q(R)}) &= \dim(Q(R) \bigotimes_R M) \\ &= \dim(Q(R)^k) + \sum \dim\left(Q(R) \bigotimes_R R/\langle p_i^{s_i} \rangle\right) \\ &= k\end{aligned}$$

Here the last line follows because $Q(R) \bigotimes_R R/\langle b \rangle = 0$ for any $0 \neq b \in R$. Indeed, a basic tensor is $a \otimes [r]_b = \frac{a}{b} \otimes [br]_b = \frac{a}{b} \otimes [0]_b = 0$. Thus k was uniquely determined. So we only need to show that we can recover the p_i and s_i's.

Next, let $q \in R$ be a prime. Then since R is a PID $R/\langle q \rangle$ is a field. Note that $R \bigotimes_R R/\langle q \rangle = R/\langle q \rangle$. Moreover, $R/\langle a \rangle \bigotimes_R R/\langle b \rangle \cong R/\langle \gcd(a, b) \rangle$, so if $q \neq p_i$ then $R/\langle q \rangle \bigotimes_R R/\langle p_i^{s_i} \rangle = 0$. If $q = p_i$, then we get $R/\langle q \rangle$, and the dimension is 1. Here really equality is up to associates. So we have:

$$\dim(M_{R/\langle q \rangle}) = k + |\{i : p_i, q \text{ are associates}\}|$$

In the above we did not get exactly what we wanted, but now we will. Consider the kernel of the map $m \mapsto p^s m$ for some s. This kernel is a submodule of M, call it K. We could compute $\dim_{R/\langle p \rangle} K_{R/\langle p \rangle}$, and hope to recover s_i. But if one tries this, it is quickly seen to fail.

Consider instead $I = \operatorname{im}(m \mapsto p^s m)$, a submodule of M. We compute $\dim_{R/\langle p \rangle}(I_{R/\langle p \rangle})$. There are several cases to handle. The places we can be mapping from are:

- R
- $R/\langle q^t \rangle$ for some q not an associate to p
- $R/\langle p^t \rangle$ where $s \geq t$

- $R/\langle p^t\rangle$ where $s < t$

For each possibility, we need to find the image of multiplication by p^s. The answers are:

- $R \mapsto \langle p^s\rangle = p^s R \cong R$ (as an R-module).
- $R/\langle q^t\rangle \mapsto R/\langle q^t\rangle$, since p^s is invertible mod q^t, since $\gcd(q,p) = 1$.
- $R/\langle p^t\rangle \mapsto 0$, since $t \leq s$
- $R/\langle p^t\rangle \mapsto \langle p^s\rangle/\langle p^t\rangle$, since $s < t$

The next thing to do is take each of these and tensor with $R/\langle p\rangle$. We get:

- $R/\langle p\rangle$
- 0, by the fact $R/\langle a\rangle \bigotimes R/\langle b\rangle \cong R/\langle \gcd(a,b)\rangle$
- 0
- $R/\langle p\rangle$

Now we count the dimensions over $R/\langle p\rangle$. They are:

- 1
- 0
- 0
- 1

So we have shown that $\dim_{R/\langle p\rangle}(\mathrm{im}(m \mapsto p^s m)) = k + |\{i : p_i, p \text{ are associates and } s_i > s\}|$. From above we already know k, so for each prime we know in how many factors it appears with power greater than $s = 1, 2, 3, \ldots$. This is enough to recover the s_i. □

We are now done with the main theorem. We have two corollaries to go.

Corollary 3.34. *Let A be a finitely generated abelian group. Then we can uniquely write $A \cong \mathbb{Z}^k \oplus \bigoplus \mathbb{Z}/\langle p_i^{s_i}\rangle \cong \mathbb{Z}^k \bigoplus \mathbb{Z}/\langle a_1\rangle \bigoplus \cdots \bigoplus \mathbb{Z}/\langle a_n\rangle$. In the first expression each p_i is prime, and in the second expression $a_1 \mid \cdots \mid a_n$.*

Proof. Any abelian group is a $\mathbb{Z}$-module. To get between the two forms, use that, for example, $\mathbb{Z}/\langle 12\rangle = \mathbb{Z}/\langle 2^2\rangle \bigoplus \mathbb{Z}/\langle 3\rangle$. One can also go the other way, but this is left as an exercise. The first expression is unique up to ordering (assuming primes to be positive), while the second is actually unique, which is part of the exercise. □

Corollary 3.35. *If F is a finite field then F^* is a cyclic group.*

Example 3.36. Let $F = \mathbb{Z}/\langle 17\rangle$. Then $|F^*| = 16$. We know that F^* is abelian, so by Corollary 3.34 we get that F^* is one of $\mathbb{Z}/\langle 16\rangle, \mathbb{Z}/\langle 8\rangle \times \mathbb{Z}/\langle 2\rangle, \ldots$. Corollary 3.35 will tell us that $F^* \cong \mathbb{Z}/\langle 16\rangle$.

Proof of Corollary 3.35. Write $F^* = \mathbb{Z}/\langle a_1\rangle \bigoplus \mathbb{Z}/\langle a_2\rangle \bigoplus \mathbb{Z}/\langle a_n\rangle$ with $a_1 \mid \cdots \mid a_n$, as in Corollary 3.34. Consider the roots of the polynomial $X^{a_1} - 1$ in F. By something we will see later, this polynomial has at most a_1 roots, since it has degree a_1. On the other hand, □

3.1. Jordan Canonical Form.

Let V be a finite dimensional vector space over a field F. Let $T : V \to V$ be a linear transformation. Then we can make V into an $F[t]$ module by $(\sum a_i t^i)v := \sum a_i T^i v$. This module is finitely generated (by any basis of V, though in general even fewer elements will suffice to generate it as an $F[t]$ module). So by our big theorem, V is isomorphic (as an $F[t]$ module) to $(F[t])^k \oplus \bigoplus F[t]/\langle p_i^{s_i}\rangle$ where p_i's are primes. Note that this isomorphism is, in particular, an isomorphism as an F-module, so we must have that $k = 0$ since $F[t]$ is an infinite-dimensional vector space over F. So we have $V \cong \bigoplus F[t]/\langle p_i^{s_i} \rangle$.

Assume that F is algebraically closed. In such a field every $f \in F[t]$ has a root $\lambda \in F$, and then $t - \lambda \mid f$, by the division algorithm in $F[t]$. So we see that if the degree of f is not 1 then f is not prime. On the other hand, it is easy to check that all elements of the form $t - \lambda$ is prime, as are their unit multiplies (where a unit in $F[t]$ is just a scalar, by a result on a homework assignment). So we have

$$V \cong \bigoplus F[t]/\langle (t - \lambda_i)^{s_i} \rangle$$

As a vector space over F, we have that $\dim(F[t]/\langle (t-\lambda_i)^s \rangle = s$, since the remainder of any polynomial when divided by $t-\lambda_i$ is of degree at most $s-1$. One basis for this vector space is $\{1, t, t^2, \dots, t^{s-1}\}$, but a nicer basis is $\{1, t - \lambda, (t - \lambda)^2, \dots, (t - \lambda)^{s-1}\}$. Let us define $e_i = (t - \lambda)^i$ for $i = 0, \dots, s - 1$, and for convenience let $e_s = 0$. Then $(t - \lambda)e_i = e_{i+1}$, so the matrix representation of $[t - \lambda]_{(e_i)}$ in this basis is

$$[t - \lambda] = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

So

$$\begin{aligned} [t] &= [t - \lambda] + \lambda I \\ &= \begin{pmatrix} \lambda & 0 & 0 \\ 1 & \lambda & 0 \\ 0 & 1 & \lambda \end{pmatrix} \end{aligned}$$

Thus we have found a basis (namely, the union of the bases we just found for the factors) such that T has a matrix representation which is block diagonal with blocks B_{λ_i, s_i}. This is the **Jordan Canonical Form**. It is canonical, meaning that it is unique up to permuting the blocks. This is because we can essentially do the above argument backwards, and then at the end apply the uniqueness of the original decomposition of V.

How does this work in practice? Let $T = \begin{pmatrix} 3/2 & 1/2 \\ 1/2 & 3/2 \end{pmatrix}$. We want to recover the matrix used in the proof of the structure theorem. It is a matrix $A \in M(F[t])$. The way we found this originally was to find some list of generators (not necessarily minimal), so we can take the generators to be the standard basis vectors for $\mathbb{R}^2$, call then v_1, v_2. Then $tv_1 = Tv_1 = \frac{3}{2}v_1 + \frac{1}{2}v_2$. The corresponding relation is $(\frac{3}{2} - t)v_1 + \frac{1}{2}v_2 = 0$. So the first column of A is $\begin{pmatrix} \frac{3}{2} - t \\ \frac{1}{2} \end{pmatrix}$. Note that this is the same as the first column of $T - tI$. Similarly, we get

$$A = \begin{pmatrix} \frac{3}{2} - t & \frac{1}{2} \\ \frac{1}{2} & \frac{3}{2} - t \end{pmatrix} = T - tI$$

How do we know that these are all of the relations? Given any $f \in F[t]$ we can write $f(t)v_i = \sum a_j v_j$ with each $a_j \in F$. We wanted to find the kernel of the map $R^2 \to V$, where $R^2 = \langle f(t)v_1 + g(t)v_2 \rangle$. This map is always surjective, and we found a one-sided inverse $v_i \mapsto 1v_i$. More work is required to see exactly why what we have done suffices.

The next thing to do is to row and column reduce the matrix A. We have:

$$\begin{aligned}\begin{pmatrix} \frac{3}{2}-t & \frac{1}{2} \\ \frac{1}{2} & \frac{3}{2}-t \end{pmatrix} &\to \begin{pmatrix} 1 & 3-2t \\ 3-2t & 1 \end{pmatrix} \\ &\to \begin{pmatrix} 1 & 0 \\ 3-2t & 1-(3-2t)^2 \end{pmatrix} \\ &\to \begin{pmatrix} 1 & 0 \\ 0 & 4t^2-12t+8 \end{pmatrix} \\ &\to \begin{pmatrix} 1 & 0 \\ 0 & t^2-3t+2 \end{pmatrix}\end{aligned}$$

We conclude that the $F[t]$ module V is isomorphic to $F[t]/\langle 1\rangle \oplus F[t]/\langle t^2-3t+2\rangle = F[t]/\langle t^2-3t+2\rangle$. Since F is algebraically closed, t^2-3t+2 is not prime. It factors as $t^2-3t+2 = (t-2)(t-1)$. Thus $V \cong F[t]/\langle (t-1)(t-2)\rangle \cong F[t]/\langle t-1\rangle \oplus F[t]/\langle t-2\rangle$. Thus the Jordan form of the original matrix T is

$$[T] = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

To actually find the basis in which this is the form of T we would need to explicitly write down the isomorphism, tracing how we changed the basis by doing row and column operations. It is also worth thinking about why this process is really the same process as the diagonalization process learning in first year linear algebra.

A crucial piece of the process is to find a polynomial to play the role of t^2-3t+2 in the above example. The necessary result is the following:

Theorem 3.37 (Cayley-Hamilton)**.** *Let A be any $n \times n$ matrix (over any commutative ring R), and define $\chi_A(t) = \det(tI - A) \in R[t]$. Then $\chi_A(A) = 0$.*

Proof. Omitted. A tempting proof is to write $\chi_A(A) = \det(AI - A) = \det(0) = 0$. This argument is, of course, nonsense. For example, $\det(0) = 0$ is a scalar, while $\chi_A(A)$ is a matrix. □

4. Unboxing Day

Recall that the goal is to unpack the Jordan Canonical Form, and see how it yields diagonalization. We will not succeed.

4.1. **Diagonalization.** Recall that given $A \in M_{n\times n}(F)$, to diagonalize one finds the eigenvalues $\lambda_1, \dots, \lambda_n$, corresponding eigenvectors $v_1, \dots, v_n$, and sets $C^{-1} = (v_1|\cdots|v_n)$, then $CAC^{-1} = D = \mathrm{diag}(\lambda_i)$.

4.2. **Jordan Form.** Again we fix $AinM_{n\times n}(F)$. $F^n =: V$ is an R-module, $R = F[t]$, by setting $tu = Au$. The generators for V are the standard basis vectors e_i. So we have a map $\pi : R^n \to V$ given by $e_i \mapsto e_i$. So $t^k e_i = A^k e_i$. This is a module morphism. We need a list of relations for $\ker \pi$. Doing this is the same as finding a map $M : R^n \to R^n$ such that the images of the standard basis under M precisely span $\ker \pi$. (We will return later to why the power of R is the same on the domain and codomain of M). Let $P_i : R^n \to R^n$ be an invertible matrix and let $Q_i : R^n \to R^n$ be invertible as well (see them in detail later). Let $M_i = Q_i^{-1} M P_i$. Then the lower row of the diagram is isomorphic to the upper row (see handout).

The challenge is to find such P, Q such that the matrix of relations is as simple as possible. If $V \cong R/\langle p_1\rangle \oplus \cdots \oplus R/\langle p_n\rangle = R/\langle t-\lambda_1\rangle \oplus \cdots \oplus R/\langle t-\lambda_n\rangle$ then the matrix of relations is $\begin{pmatrix} t-\lambda_1 & \cdots & 0 \\ 0 & \cdots & 0 \\ 0 & \cdots & t-\lambda_n \end{pmatrix}$. Call this matrix M_2, then $V \cong R^n/\operatorname{im}(M_2)$. This is clear by thinking of how M_2 acts on column vectors from R^n. Recall that if the a_i's are relatively prime then $\bigoplus R/\langle a_i\rangle \cong R/\langle \prod a_i\rangle$. The matrix giving this is diagonal with 1's until the last place, which has $\prod a_i$ in the bottom right.

In the diagram on the handout, we claim that the matrix M of relations is given by $M = tI - A$.

Proof. Expand A as $A = (a_{ij})$. We first claim that anything of the form $te_i - Ae_i$ is a relation (where e_i is a standard basis vector of F). That is, we are claiming that $\pi(te_i - Ae_i) = 0$. But this is clear, since $\pi(te_i) = Ae_i$, while $\pi(Ae_i) = Ae_i$.

Now we claim that these are all of the relations. That is, we are claiming that if $\pi(u) = 0$, then u is an R-linear combination of the $r_i := te_i - Ae_i$. This is the same as showing that $\operatorname{im}(tI - A) = \{\text{relations}\}$. Let $\iota : V \to R^n$ be the map sending the ith basis vector of V to e_i. ι is F-linear, but may not be $R = F[t]$-linear. For any u, we will show that $u - \iota(\pi(u))$ is a linear combination of the relations r_i. It suffices to do this for the case $u = t^k e_i$. Then

$$\begin{aligned} u - \iota(\pi(u)) &= t^k e_i - A^k e_i \\ &= (t^k I - A^k)e_i \\ &= (tI - A)(t^{k-1}I + t^{k-2}A + \cdots + t^0 A^{k-1})e_i \\ &\in \operatorname{im}(tI - A) \end{aligned}$$

Now we see that if $\pi(u) = 0$ then u itself will be a combination of the r_i's, as claimed. □

Now we need to know how P_2, Q_2 are related to M_2, C where C is the matrix from undergraduate diagonalization. One direction is easy, given C take $P_2 = Q_2^{-1} = C$, and see the handout. Now suppose that we are given P_2, Q_2. If it turns out that P_2, Q_2 are matrices of scalars, then one easily checks that $P_2 = Q_2^{-1}$, and we recover C. But there is no reason to believe this will be the case.

We already have a map $\iota : V \to R^n$, so $C = \pi' P_2 \iota$. $Ce_i = \pi'(P_2 e_i)$. Write $P_2 e_i = \sum t^k u_k$ where $u_k \in F^n$, then $\pi'(P_2 e_i) = \sum D^k u_k$. Put another way, write $P_2 = \sum t^k P_{2,k}$ where $P_{2,k} \in M_{n\times n}(F)$. Then $C = \sum D^k P_{2,k}$. That is, $C = L_D P_2$, where L_D is left-evaluation at D.

University of Toronto
E-mail address: cjeagle@math.utoronto.ca