Lethbridge Advanced Abstract Algebra

Dave Witte Morris

University of Lethbridge, Canada

December 22, 2019

Contents

Contents

Contents

Part I

Group Theory

Chapter 1

Summary of undergraduate group theory

Although this course assumes familiarity with the topics in a typical undergraduate course on abstract algebra, including subgroups, normal subgroups, homomorphisms, quotient groups, etc., we will start with a quick review.

(1.0.1) **Notation.**

- G is always a group, and X is a set.
- The cardinality of X is denoted $|X|$ (or, sometimes, $\#X$). (Recall that the ***cardinality*** of a set is the number of elements in the set.)
- $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$.
- $\mathbb{N} = \{0, 1, 2, \ldots\}$ (unlike some other authors, we include 0 in this set).
- $\mathbb{Z}^+ = \mathbb{N}^+ = \{1, 2, 3, \ldots\}$.
- For $k, n \in \mathbb{Z}$, we write $k \mid n$ to denote that k is a divisor of n (or, equivalently, that n is a multiple of k).
- For $a, b, n \in \mathbb{Z}$, we write $a \equiv b \pmod{n}$ to denote that $n \mid (a - b)$, and we may say that a is congruent to b, modulo n.

§1.1. Definitions and examples

(1.1.1) **Definition.** A ***group*** is a set G together with a binary operation $*$ that is associative and has an identity element and inverses. (We usually write gh, or sometimes $g \cdot h$, for $g * h$.)

- (associative) $\forall g_1, g_2, g_3 \in G,\ g_1(g_2g_3) = (g_1g_2)g_3$.
- (identity element) $\exists e \in G, \forall g \in G,\ eg = ge = g$.

  It is easy to show that the identity element of G is unique (*see Example 1.1.2*(1)). *It is usually denoted by* 1 *or, sometimes, e.* (*However, it is denoted by* 0 *if G is written additively, which means that the group operation is* $+$.) *To avoid confusion, we will sometimes use 1_G for the identity element of G* (*and 1_H for the identity element of some other group H*).
- (inverses) $\forall g \in G, \exists h \in G,\ gh = hg = 1$.

  It is easy to show that the inverse of g is unique (*see Example 1.1.2*(2)). *It is denoted g^{-1}* (*unless G is written additively, in which case the inverse is $-g$*) .

(1.1.2) **Example.** We verify two facts stated in Definition 1.1.1, and also establish two facts of high-school algebra.

1) If e_1 and e_2 are identity elements of G, then $e_1 = e_2$. To see this, note that:
$$\begin{aligned} e_1 &= e_1 e_2 && (e_2 \text{ is an identity element, so } g e_2 = g \text{ for all } g \in G) \\ &= e_2 && (e_1 \text{ is an identity element, so } e_1 g = g \text{ for all } g \in G). \end{aligned}$$
2) Assume that G has an identity element 1, and let $g \in G$. If h_1 and h_2 are inverses of g, then $h_1 = h_2$. To see this, note that:
$$\begin{aligned} h_1 &= h_1 * 1 && (1 \text{ is the identity element, so } h * 1 = h \text{ for all } h \in G) \\ &= h_1 * (g * h_2) && (h_2 \text{ is an inverse of } g) \\ &= (h_1 * g) * h_2 && (* \text{ is associative}) \\ &= 1 * h_2 && (h_1 \text{ is an inverse of } g) \\ &= h_2 && (1 \text{ is the identity element, so } 1 * h = h \text{ for all } h \in G). \end{aligned}$$
3) The inverse of a product is the product of the inverses **in the reverse order**:
$$(gh)^{-1} = h^{-1} g^{-1}.$$
To see this, note that
$$(gh)(h^{-1}g^{-1}) = g(hh^{-1})g^{-1} = g \cdot 1 \cdot g^{-1} = gg^{-1} = 1.$$
A similar calculation shows that $(h^{-1}g^{-1})(gh) = 1$. Thus, we have shown that if you multiply gh by $h^{-1}g^{-1}$ on either side, then the result is the identity element. This is exactly what it means to say that $h^{-1}g^{-1}$ is the inverse of gh.
4) G has both right and left cancellation: if $ag = ah$ or $ga = ha$, then $g = h$. To see this, note that if $ag = ah$, then $a^{-1}(ag) = a^{-1}(ah)$. Furthermore (using the associative law) the left-hand side is $1 \cdot g = g$ and the right-hand side is $1 \cdot h = h$, so we conclude that $g = h$, as desired. Similarly, if $ga = ha$, then
$$g = g \cdot 1 = (ga)a^{-1} = (ha)a^{-1} = h \cdot 1 = h.$$

(1.1.3) **Definitions.** Let g, h, and a be elements of a group G.

1) The cardinality of the set G is called the ***order*** of G. It is denoted $|G|$.
2) For $k \in \mathbb{Z}^+$, we define g^k to be the product of k copes of g: we have $g^k = gg \cdots g$, where there are k factors on the right-hand side. And we let $g^{-k} = (g^k)^{-1}$ (or $(g^{-1})^k$, which is the same thing, by Example 1.1.2(3)). Finally, we let $g^0 = 1$. With these definitions, the usual laws of exponents hold (for $k, \ell \in \mathbb{Z}$):
$$g^0 = 1, \quad g^1 = g, \quad g^k g^\ell = g^{k+\ell}, \quad (g^k)^\ell = g^{k\ell}, \quad (g^k)^{-1} = (g^{-1})^k.$$
(If the group operation is $+$, then we write kg for $g + g + \cdots + g$, instead of g^k.)
3) The ***order*** of g is the smallest $k \in \mathbb{Z}^+$, such that $g^k = 1$. It is denoted $|g|$. (If no such k exists, then $|g| = \infty$.)
4) g and h ***commute*** if $gh = hg$. (We may also say that g ***centralizes*** h.) We say that G is ***abelian*** (or ***commutative***) if all the elements of G commute with each other.

(1.1.4) **Exercise.** Let $g \in G$, such that g has finite order.

1) For $k, \ell \in \mathbb{Z}$, show:
   (a) $g^k = 1$ if and only if k is a multiple of $|g|$.
   (b) More generally, $g^k = g^\ell \Leftrightarrow k \equiv \ell \pmod{|g|}$.
2) Show $|g^{-1}| = |g|$.

In this course, we will mostly be interested in **finite** groups. (These are groups that have only finitely many elements, or in other words, the groups whose order is finite.) It is important to be familiar with some examples.

(1.1.5) **Example** (Integers modulo n under addition)**.** Recall that $\mathbb{Z}_n$ is the set of ***integers modulo n***. This means that elements of $\mathbb{Z}_n$ are integers, except that we consider two elements of $\mathbb{Z}_n$ to be equal if they are congruent modulo n. More precisely, the elements of $\mathbb{Z}_n$ are *equivalence classes* of integers, where two integers are equivalent if they have the same remainder when you divide them by n. We can use $\overline{k}$ to represent the equivalence class of k in $\mathbb{Z}_n$. For example,

$$\mathbb{Z}_2 = \{\overline{0}, \overline{1}\}, \mathbb{Z}_3 = \{\overline{0}, \overline{1}, \overline{2}\}, \mathbb{Z}_4 = \{\overline{0}, \overline{1}, \overline{2}, \overline{3}\}, \mathbb{Z}_5 = \{\overline{0}, \overline{1}, \overline{2}, \overline{3}, \overline{4}\}, \mathbb{Z}_6 = \{\overline{0}, \overline{1}, \overline{2}, \overline{3}, \overline{4}, \overline{5}\}, \text{ etc.}$$

Each of these is a group under addition modulo n, which is defined by $\overline{k} + \overline{\ell} = \overline{k + \ell}$:

1) (associative) In elementary school, we all learned the associative law (for the addition of ordinary integers). This implies that addition modulo n is also associative:

$$\begin{aligned}
(\overline{k} + \overline{\ell}) + \overline{m} &= \overline{k + \ell} + \overline{m} && \text{(definition of addition in } \mathbb{Z}_n) \\
&= \overline{(k + \ell) + m} && \text{(definition of addition in } \mathbb{Z}_n) \\
&= \overline{k + (\ell + m)} && \text{(associate law for addition in } \mathbb{Z}) \\
&= \overline{k} + \overline{\ell + m} && \text{(definition of addition in } \mathbb{Z}_n) \\
&= \overline{k} + \overline{\ell} + \overline{m} && \text{(definition of addition in } \mathbb{Z}_n).
\end{aligned}$$

2) (identity element) The additive identity element of $\mathbb{Z}$ is 0, so the identity element of $\mathbb{Z}_n$ is $\overline{0}$:
$$\overline{k} + \overline{0} = \overline{k + 0} = \overline{k} \quad \text{and} \quad \overline{0} + \overline{k} = \overline{0 + k} = \overline{k}.$$

3) (inverses) The (additive) inverse of k in $\mathbb{Z}$ is $-k$ so the inverse of $\overline{k}$ in $\mathbb{Z}_n$ is $\overline{-k}$:
$$\overline{k} + \overline{-k} = \overline{k + (-k)} = \overline{0} \quad \text{and} \quad \overline{-k} + \overline{k} = \overline{(-k) + k} = \overline{0}.$$

Note that, for the element $\overline{1} \in \mathbb{Z}_n$, we have $|\overline{1}| = n$.

(1.1.6) **Exercise.** The ***direct product*** of two groups G and H is the Cartesian product $G \times H$ with componentwise multiplication. That is, $(g_1, h_1) * (g_2, h_2) = (g_1 g_2, h_1 h_2)$.

1) Show that $G \times H$ is a group.
[*Hint:* By definition, this means you need to show that the operation is associative, and has an identity element and inverses.]

2) For $(g, h) \in G \times H$, show that $|(g, h)|$ is the least common multiple of $|g|$ and $|h|$.

(1.1.7) **Example.** $\mathbb{Z}_3 \times \mathbb{Z}_5 \times \mathbb{Z}_7$ is a group of order $3 \cdot 5 \cdot 7 = 105$. In this group, we have
$(\overline{2}, \overline{4}, \overline{5}) + (\overline{2}, \overline{1}, \overline{4}) = (\overline{2 + 2}, \overline{4 + 1}, \overline{5 + 4}) = (\overline{2 + 2}, \overline{4 + 1}, \overline{5 + 4}) = (\overline{4}, \overline{5}, \overline{9}) = (\overline{1}, \overline{0}, \overline{2})$.

(1.1.8) **Exercise** (semidirect product of cyclic groups)**.** Suppose $m, n \in \mathbb{Z}^+$, and let $k \in \mathbb{Z}$, such that $k^n \equiv 1 \pmod{m}$. Define a binary operation on the set $\mathbb{Z}_m \times \mathbb{Z}_n$ by

$$(x_1, y_1) * (x_2, y_2) = (x_1 + k^{y_1} x_2, y_1 + y_2).$$

Show that this operation is associative, and has an identity element and inverses. Hence, it defines a group (of order mn).

This group is called the ***semidirect product*** of $\mathbb{Z}_m$ and $\mathbb{Z}_n$ (with multiplier k), and is denoted $\mathbb{Z}_m \rtimes_k \mathbb{Z}_n$. It is a generalization of the the direct product of $\mathbb{Z}_m$ and $\mathbb{Z}_n$, which is the special case where $k = 1$. (The construction can be generalized by replacing $\mathbb{Z}_m$ and $\mathbb{Z}_n$ with other groups, but we will not discuss this.)

(1.1.9) **Example.** In $\mathbb{Z}_7 \times_4 \mathbb{Z}_3$, we have

$$(\overline{3}, \overline{2}) * (\overline{5}, \overline{1}) = (\overline{3 + 4^2 \cdot 5},\ \overline{2 + 1}) = (\overline{3 + 4^2 \cdot 5},\ \overline{2 + 1}) = (\overline{83},\ \overline{3}) = (\overline{6},\ \overline{0}).$$

Groups often arise as symmetries of an object.

(1.1.10) **Definition** (informal)**.** A ***symmetry*** of an object is a way of repositioning the object in such a way that it occupies exactly the same space as it did originally. It is sometimes called an "undetectable motion:" if you perform a symmetry to an object while someone is not looking, they will not realize that anything has changed, because everything looks exactly as it did before.

(1.1.11) **Example** (Rotations of a square)**.** Imagine a square lying on a tabletop. Rotating the square by 90° is a symmetry of the square. For short, let us use r_θ to denote a rotation by θ degrees (clockwise). So r_{90} is a symmetry of the square. Other symmetries are r_{180} and r_{270}. (By the way, another name for r_{270} is r_{-90}. Or we could rotate by 0° (doing nothing certainly leaves the square occupying the same space as it did before), so $r_0 = r_{360}$ is also a symmetry of the square. There are no other rotational symmetries, so the rotational symmetries of the square form the set $\{r_0, r_{90}, r_{180}, r_{270}\}$.

It is important to note that this set is a group under composition, or, in the language of non-mathematicians, the "after" operation: recall that $r \circ s$ is the procedure that is obtained by doing r after s: first apply s to the object, then apply s. For example, $r_{90} \circ r_{180} = r_{270}$, because rotating the square by 90 degrees after rotating it by 180 degrees has exactly the same effect as applying a single rotation of 180 degrees to the square. Note that this is a group of order 4. We have $|r_0| = 1$, $|r_{90}| = |r_{270}| = 4$, and $|r_{180}| = 2$.

In general, the symmetries of an object always form a group.

(1.1.12) *Remark.* It is not difficult to verify the above statement that symmetries form a group:

- To see that composition is a binary operation on the set of symmetries, we need to know that the set of symmetries is closed under composition: if r and s are symmetries of an object X, then $r \circ s$ is also a symmetry of X. Fortunately, this is clearly true: if we reposition X in some way that is undetectable, and then reposition it again in a way that is undetectable, then the final position is also indistinguishable from the original position.
- Composition is associative: $r \circ (s \circ t) = (r \circ s) \circ t$. This is because it does not matter whether we :
  - first do t, then do s, and then take a break before doing r, or
  - first do t, then take a break before doing s and then r.

  In either case, we are doing t, and then s, and then r.
- The identity element of the group is the "do nothing" operation.
- The inverse of a particular symmetry is "put it back the way it was."

The square can be replaced with any regular polygon:

(1.1.13) **Example.** The set of rotations of a regular n-gon P is a group $\mathrm{Rot}(P)$ of order n. More precisely, if $\theta = 360/n$, then

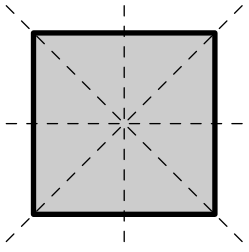
$$\mathrm{Rot}(P) = \{r_0, r_\theta, r_{2\theta}, \ldots, r_{(n-1)\theta}\}.$$

Furthermore, we have $r_{k\theta} \circ r_{\ell\theta} = r_{(k+\ell)\theta}$, and $r_{k\theta}^{-1} = r_{-k\theta}$.

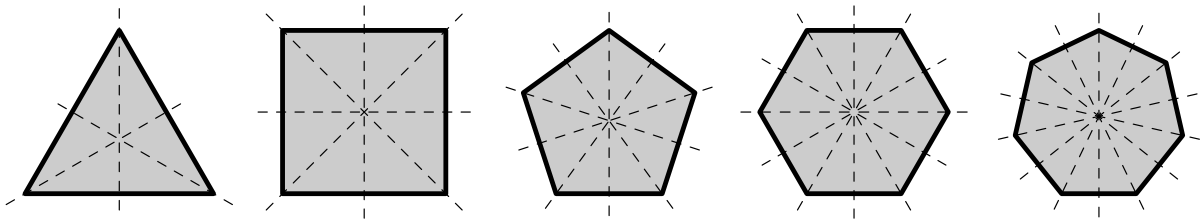
(1.1.14) **Exercise.** Find the order of each of the six rotations of the regular hexagon.

But rotations are not the only symmetries of a regular n-gon:

(1.1.15) **Example** (dihedral group)**.** In addition to rotational symmetry, a regular n-gon also has reflection symmetry: it looks the same in a mirror. Another way of looking at this is that the square can be "flipped over." For example, if S is a square in the xy-plane, with its centre at the origin and its sides parallel to the coordinate axes, then rotating S by 180° about the x-axis is a symmetry of S. (Another way of describing this is to perform a mirror symmetry across the x-axis: we are simply moving each point (x, y) to $(x, -y)$.) The y-axis is also an axis of symmetry. And there are two other reflection axes that go through opposite corners of the square. They are illustrated in the figure at right.



In general, a regular n-gon has precisely n reflection symmetries: the axes are lines through the origin whose intersections with the boundary of the n-gon are either at a corner or are the exact midpoint of an edge.



Thus, the full group of symmetries of a regular n-gon has order $2n$: there are n rotations and n reflections. (Every reflection has order 2, but the orders of the rotations are various numbers between 1 and n, inclusive.) This group is called the ***dihedral group*** of order $2n$, and is denoted D_{2n}. For example, the group of symmetries of a square is D_8, and the group of symmetries of a regular pentagon is D_{10}.

(1.1.16) *Other conventions.* Some textbooks use the notation D_n for the dihedral group of order $2n$, instead of D_{2n}.

We can also look at objects in 3-dimensions, instead of the plane:

(1.1.17) **Example** (rotations of a cube)**.** A cube has 24 rotational symmetries. To see this, consider the cube to being resting on a tabletop. We can move any of the 6 faces of the cube to be on the bottom. After this, the face that is now on the bottom can be in any of 4 positions (because the face is a square, and a square has 4 rotational symmetries). Therefore, we have 6 choices for which face to put on the bottom, and 4 choices of position after that. (And these two choices completely determine the position of the cube.) So the total number of possible positions (or symmetries) is $6 \times 4 = 24$.

(1.1.18) **Exercises.**

1) Recall that the Platonic solids are:
   (a) regular tetrahedron (the 4 faces are equilateral triangles),
   (b) cube (the 6 faces are squares),
   (c) regular octahedron (the 8 faces are equilateral triangles),
   (d) regular dodecahedron (the 12 faces are regular pentagons), and
   (e) regular icosahedron (the 20 faces are equilateral triangles).

   What is the order of the group of rotations of each of these solids?

   [*Hint:* Let us say that the solid is resting on a tabletop. You may assume, without proof, that any face can be rotated to be on the bottom, that each rotational symmetry of the bottom face extends to a symmetry of the entire solid, and that these two choices determine the symmetry of the solid. (This is because we are dealing only with Platonic solids.)]
2) How many symmetries does a rectangle have if it is *not* a square? (Include both rotations and reflections.)
3) Describe each of the rotational symmetries of a regular tetrahedron. In particular, how many are there of each order?
4) Describe each of the 24 rotational symmetries of a cube. In particular, how many are there of each order?

We can also look at symmetries of a collection of objects, instead of a single object. For example, if there are several identical items sitting on a table, then we could interchange some of them, and the scene would look exactly the same as before. This is an example of a permutation. To keep track of what we are doing, it is helpful to number the items, so that we can say, for example, that we interchanged item #2 with item #6.

(1.1.19) **Definition.**

1) Recall that a function $f\colon X \to Y$ is:
   - ***one-to-one*** (or ***injective***) iff for all $x_1, x_2 \in X$, such that $f(x_1) = f(x_2)$, we have $x_1 = x_2$;

- ***onto*** (or ***surjective***) iff for all $y \in Y$, there exists $x \in X$, such that $f(x) = y$;
- a ***bijection*** iff f is both one-to-one and onto.

2) A ***permutation*** of X is a bijection from X to itself.
3) The set of all permutations of X is called the ***symmetric group*** on X. It is a group under composition (see Exercise 1.1.24): $(\sigma\tau)(x) = \sigma(\tau(x))$ for all $x \in X$.
4) The symmetric group on $\{1, 2, \ldots, n\}$ is called the ***symmetric group on n letters*** (or "of degree n"), and is denoted S_n. Its order is $n!$.
5) For $x_1, \ldots, x_k \in X$, we use $(x_1\, x_2\, \ldots\, x_k)$ to denote the unique permutation $\sigma \in S_n$, such that
   - $\sigma(x_i) = x_{i+1}$ for $i \in \{1, \ldots, k\}$ (reading the subscript modulo k), and
   - $\sigma(x) = x$ for all $x \notin \{x_1, x_2, \ldots, x_k\}$.

   Such a permutation is called a ***cycle*** of length k, or a ***k-cycle***.
6) Two cycles $(x_1\, x_2\, \ldots\, x_k)$ and $(y_1\, y_2\, \ldots\, y_\ell)$ are ***disjoint*** if the sets $\{x_1, x_2, \ldots, x_k\}$ and $\{y_1, y_2, \ldots, y_\ell\}$ are disjoint (that is, they have no elements in common).

Dealing with permutations requires some basic facts from Math 2000. If you have difficulty with the definition, or the subsequent exercises, you may want to review a textbook for that course. One such textbook is available online at:

`http://people.uleth.ca/~dave.morris/books/proofs+concepts.html`

(1.1.20) **Definition.** Assume $\varphi\colon X \to Y$ and $\sigma\colon Y \to Z$. The ***composition*** of σ and φ is the function $\sigma \circ \varphi$ from X to Z that is defined by

$$(\sigma \circ \varphi)(x) = \sigma(\varphi(x)) \quad \text{for all } x \in X.$$

(1.1.21) **Exercises** (Basic properties of composition)**.** Assume $\varphi\colon X \to Y$ and $\sigma\colon Y \to Z$.

1) Prove that composition is associative: show that if, in addition to $\varphi\colon X \to Y$ and $\sigma\colon Y \to Z$, we also have $\tau\colon Z \to W$, then $(\tau \circ \sigma) \circ \varphi = \tau \circ (\sigma \circ \varphi)$. (This was explained informally in Remark 1.1.12, but you should be able to write an official proof of this fact.)
2) Show that the composition of one-to-one functions is one-to-one: if φ and σ are one-to-one, then $\sigma \circ \varphi$ is one-to-one.
3) Show that the composition of onto functions is onto: if φ and σ are onto, then $\sigma \circ \varphi$ is onto.
4) Show that the composition of bijections is a bijection: if φ and σ are bijections, then $\sigma \circ \varphi$ is a bijection.

   [*Hint:* Use (2) and (3).]

Bijections arise in many situations, but one of their most important applications is in showing that two sets have the same cardinality:

(1.1.22) **Basic facts.**

1) **Definition.** Two sets X and Y have the ***same cardinality*** if and only if there is a bijection from X to Y.
2) Assume $\varphi\colon X \to Y$, and let A be any subset of X.
   (a) **Definition.** $f(A) = \{\, f(a) \mid a \in A \,\}$. (This is called the ***image*** of A under f.)
   (b) If φ is one-to-one, then $|f(A)| = |A|$.

It is not difficult to see if $\varphi\colon X \to Y$ has an inverse $\varphi^{-1}\colon Y \to X$, then φ must be a bijection. It is an important fact from Math 2000 that the converse is true. However, this is more difficult, so you do not need to try to prove it for yourself, although you should remember this important fact:

(1.1.23) **Basic fact.** If $\varphi\colon X \to Y$ is a bijection, then φ has an inverse $\varphi^{-1}\colon Y \to X$.

Some of the above facts will be helpful in solving the following problem:

(1.1.24) **Exercise.** Prove that the symmetric group S_X is indeed a group (under composition).

Also show that the identity element of this group is the identity map on X. (The ***identity map*** on X is the function $\mathbb{1}_X\colon X \to X$ defined by $\mathbb{1}_X(x) = x$ for $x \in X$.)

(1.1.25) **Basic facts.**

1) If two cycles are disjoint, then they commute with each other.
2) Every permutation of a finite set is a product of disjoint cycles. Furthermore, this decomposition into disjoint cycles is unique, up to a permutation of the factors.

(1.1.26) **Example.** $(1\,3)(2\,5\,7)$ is an element of S_7. (In fact, it is an element of S_n for any $n \geq 7$.) Its action on the elements of $\{1,2,3,4,5,6,7\}$ is:

$$1 \mapsto 3, \quad 2 \mapsto 5, \quad 3 \mapsto 1, \quad 4 \mapsto 4, \quad 5 \mapsto 7, \quad 6 \mapsto 6, \quad 7 \mapsto 2.$$

This permutation has order 6. (In general, the order of a permutation is the least common multiple of the lengths of the cycles in its decomposition as a product of disjoint cycles.)

(1.1.27) **Exercise.** In each part of the problem, write each of the given permutations as a product of disjoint cycles. Also find the order of each of the permutations.

1) $(1\,2)(2\,3)(3\,4)$. 2) $(1\,2)(1\,3)(1\,4)(1\,5)$. 3) $(2\,3\,5)(2\,3\,4)(1\,4\,2\,5\,3)$.
4) Number the corners of a regular hexagon from 1 to 6, clockwise, so that we may think of each of its six rotational symmetries as elements of S_6. You are to consider all six of these permutations.
5) Number the corners of a square from 1 to 4, clockwise, so that we may think of each of the elements of D_8 as elements of S_6. The permutations to consider are:
   (a) the reflection whose axis is through corners 1 and 3, and
   (b) the reflection whose axis is through the midpoints of sides $1-2$ and $3-4$.

§1.2. Burnside's Counting Lemma (optional)

(1.2.1) **Problem.** We have five colours of paint available with which to paint the faces of a cube. (Every face needs to be painted, and only one colour is allowed to appear on each face, but several faces may be painted the same colour, and it is not necessary to use all 5 colours.) How many essentially different ways are there to paint the cube?

High-school-level mathematics reveals there are exactly $5^6 = 15,625$ ways to paint the faces of the cube, because we can use any one of 5 colours of paint on each of the 6 faces of the cube. However, this is not the answer we want. For example, suppose Alice paints the top face of her cube blue and all of the other faces red, while Bob paints the bottom of his cube blue and all the other faces red. Then Alice and Bob have painted their cubes in essentially the same way — if Alice turns her cube upside-down, then it looks just like Bob's.

If there were only two colours of paint, it would not be difficult to solve the problem by listing all of the possible colourings. But that is not very feasible for five colours (unless a computer is used). This section presents a method that easily solves this problem and many related types of problems, by applying the theory of group actions.

We will see the official definitions in Section 2.1, but, for now, it will suffice to have an informal understanding of two key ideas:

- Saying that a group G *acts* on a set X means that each element g of G acts like a permutation of the elements of a set X: the group element g carries each $x \in X$ to some point that is called gx.
- Each x in X can be moved to certain other places in X. (But probably there are places that G cannot move it to. For example, rotations of a square cannot move a corner point to a point that is not on a corner.) The set of points that x can move to is called the *orbit* of x.

(1.2.2) **Definition.** Suppose G acts on X. For $g \in G$ and $x \in X$, we say that x is a ***fixed point*** of g if $gx = x$. (This means that g leaves x alone, instead of moving it somewhere else.)

(1.2.3) **Proposition** (Burnside's Counting Lemma)**.** *Suppose G acts on a finite set X (and G is finite). Then the number of orbits of G on X is equal to the number of fixed points of the average element of G.*

(1.2.4) *Remark.* To be precise, the phrase "the number of fixed points of the average element of G," actually means "the average of the numbers of fixed points of elements of G." That is, list the number of fixed points of each element of G, and then calculate the average (or arithmetic mean) of these numbers.

We will see some applications of this proposition right now, but the proof will be postponed until Section 2.2, when we have more tools from the theory of group actions.

(1.2.5) **Examples.**

1) Suppose every element of G leaves all of the points of X alone. (This could be called the "lazy" group action, because the elements of G are too lazy to do anything, but it is officially called the *trivial* action of G.) Then each element of X constitutes an orbit, so the number of orbits is $|X|$. On the other hand, each element of G fixes every point in X, so the number of fixed points of each element (including the average element) is also $|X|$. This agrees with the conclusion of Burnside's Counting Lemma.
2) Let $G = \mathrm{Rot}(\text{cube})$ and let X be the set of faces of the cube. Only the identity element of G fixes every face of the cube, while there are 9 rotations that fix two faces (six rotations of order 4, and three rotations of order 2), and all other elements of G have no fixed points in the action on X. So the number of fixed points of the average element of G is
$$\frac{1 \times 6 + 9 \times 2 + 14 \times 0}{24} = 1.$$
Also, any face of the cube can be moved to any other face by a rotation, so this action has only 1 orbit. Hence, we once again have agreement with the conclusion of Burnside's Counting Lemma.
3) Let G be the group of symmetries of a rectangle (which is *not* a square), and let X be the set of sides of the rectangle. The rectangle has only four symmetries (see Exercise 1.1.18(2)):
   - The trivial symmetry (do nothing) fixes all 4 sides.
   - r_{180} does not fix any sides, so the number of sides that are fixed is 0.
   - Each of the two reflection symmetries fixes the 2 sides that are on its axis (and interchanges the other two).

   So the number of fixed points of the average element of G is
$$\frac{1 \times 4 + 1 \times 0 + 2 \times 2}{4} = 2.$$
Also the long sides of the rectangle can be moved to each other, and the short sides of the rectangle can be moved to each other, but a symmetry cannot move a long side so a short side. So there are 2 orbits. Therefore, of course, we yet again have agreement with the conclusion of Burnside's Counting Lemma.

Solution to Problem 1.2.1. Two colourings of the cube are essentially different if there is no rotation of the cube that takes one to the other, that is, if they are in different orbits in the action of Rot(cube) on the collection of colourings of the cube. So the total number of essentially different colourings of the cube is equal to the number of orbits of Rot(cube) on the set of colourings. By Burnside's Counting Lemma, this is the same as the number of colourings fixed by an average element of Rot(cube).

Any permutation can be uniquely expressed as a product of disjoint cycles, and it is easy to see that such a permutation will fix a colouring if and only if each cycle of the permutation is monochromatic. (That is, all of the faces in each cycle must have the same colour, but faces in different cycles may have different colours.) Thus, if there are k colours available, then the number of colourings fixed by the permutation g is $k^{\mathrm{cyc}(g)}$, where $\mathrm{cyc}(g)$ is the number of cycles in the disjoint cycle decomposition of g. Therefore, applying Burnside's Counting Lemma tells us:

The number of essentially different ways to paint the faces of a cube with k colours is

$$\frac{1}{|\mathrm{Rot}(\mathrm{cube})|}\sum_{g\in\mathrm{Rot}(\mathrm{cube})} k^{\mathrm{cyc}(g)}.$$

To calculate the sum, we count the cycles in each element of Rot(cube):

1) The identity element fixes all 6 faces, so it has 6 cycles of length one.
2) 3 elements are rotations of 180° about an axis through opposite faces of the cube. These have 2 cycles of length one, and 2 cycles of length two, for a total of 4 cycles.
3) 6 elements are 90° rotations. They have 2 cycles of length one, and 1 cycle of length four, for a total of 3 cycles.
4) 6 elements are 180° rotations about an axis through opposite edges. They have 3 cycles, all of length two.
5) 8 elements are rotations about an axis through opposite corners. They have 2 cycles of length three.

So the number of essentially different colourings is

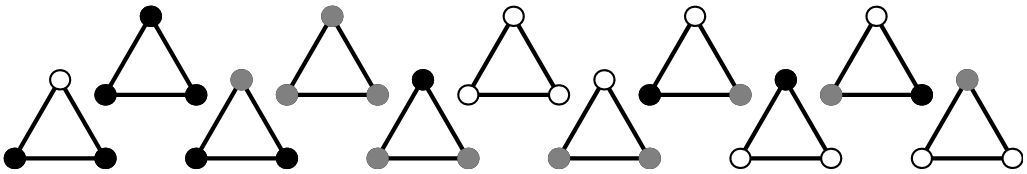
$$\frac{k^6 + 3k^4 + 6k^3 + 6k^3 + 8k^2}{24}. \tag{1.2.6}$$

We plug in $k = 5$ to get 800 as the answer to the original problem. □

(1.2.7) *Remark.* Letting $k = 1$ in (1.2.6) yields the value 1. This agrees with the observation that there is only one way to colour the faces of a cube if only one colour of paint is available. (The entire cube must be covered with that one colour.)

(1.2.8) **Exercises.**

1) Letting $k = 2$ in (1.2.6) tells us that there are exactly 10 essentially different ways to colour the faces of a cube with two colours. Describe (or draw) each of these 10 colourings.
2) How many essentially different ways are there to colour the faces of a cube with three colours?
3) How many essentially different ways are there to colour the vertices of an equilateral triangle if k colours of paint are available?
   (a) Assume two colourings are equivalent if one can be obtained from the other by rotating the triangle.
   For $k = 3$, there are 11 different colourings:



   (b) Assume two colourings are equivalent if one can be obtained from the other by applying any symmetry of the triangle (that is, either a rotation or a reflection of the triangle).
   For $k = 3$, there are 10 different colourings. They are the same colourings pictured above, except that there is now only one colouring that uses all 3 colours.
4) Replace the triangle in Exercise 3 with a regular p-gon, where p is a prime number (and $p \geq 3$). Consider both (a) rotations only and (b) all of the symmetries in the dihedral group D_{2p}.
5) For the special case of Exercise 4a with $p = 5$ and $k = 2$, draw one colouring from each of the 8 different equivalence classes.
6) Replace the cube in Problem 1.2.1 with a regular tetrahedron (and find a formula that solves this problem for any number k of colours, not only for 5 colours).

§1.3. Subgroups, conjugates, cosets, and quotient groups

(1.3.1) **Definitions** (subgroups)**.**

1) A subset H of G is a ***subgroup*** of G if it is closed under the group operations: for all $h_1, h_2, h \in H$, we have $h_1 h_2 \in H$ and $h^{-1} \in H$ (and $1 \in H$). Equivalently, this means that H is itself a group under the operation obtained by restricting the operation of G.
   (For example, if P is a regular n-gon, then $\mathrm{Rot}(P)$ is a subgroup of D_{2n}.)
2) The obvious subgroups of G are $\{1\}$ and G. (There may or may not be other subgroups of G.)
   - $\{1\}$ is the ***trivial*** subgroup of G. (All other subgroups are ***nontrivial***.)
   - A subgroup H of G is said to be ***proper*** if $H \neq G$.
3) For $g \in G$, we let $\langle g \rangle = \{ g^k \mid k \in \mathbb{Z} \}$. This is an abelian subgroup of G (see Exercises 1.3.4(1) and 1.3.5(1)), and is called the ***cyclic subgroup*** generated by g.
4) We say G is ***cyclic*** if $G = \langle g \rangle$ for some $g \in G$. For example, $\mathbb{Z}_n$ is cyclic (for every $n \in \mathbb{Z}^+$), because $\mathbb{Z}_n = \langle \overline{1} \rangle$.

(1.3.2) **Example.** If H_1 and H_2 are subgroups of G, then $H_1 \cap H_2$ is a subgroup of G.

Proof. We wish to show that $H_1 \cap H_2$ is closed under multiplication and inverses, and contains the identity element.

(closed under multiplication) Given $g, h \in H_1 \cap H_2$ we have $g, h \in H_1$ and $g, h \in H_2$. Since each H_i is a subgroup, and is therefore closed under multiplication, this implies that $gh \in H_1$ and $gh \in H_2$. So $gh \in H_1 \cap H_2$, as desired.

(closed under multiplication) Given $h \in H_1 \cap H_2$ we have $h \in H_1$ and $h \in H_2$. Since each H_i is a subgroup, and is therefore closed under inverses, this implies that $h^{-1} \in H_1$ and $h^{-1} \in H_2$. So $h^{-1} \in H_1 \cap H_2$, as desired.

(identity element) Since each H_i is a subgroup, it must contain the identity element. This means that $1 \in H_1$ and $1 \in H_2$. So $1 \in H_1 \cap H_2$, as desired. □

(1.3.3) *Remark.* By induction, Example 1.3.2 implies that the intersection of any finite collection of subgroups of G is a subgroup of G, because $H_1 \cap H_2 \cap \cdots \cap H_n = (H_1 \cap H_2 \cap \cdots \cap H_{n-1}) \cap H_n$ is the intersection of two subgroups. But it is not difficult to show directly that the intersection of any collection of subgroups is also a subgroup, even if the collection is infinite.

(1.3.4) **Exercise.** Here are some important subgroups of G:

1) Let $g \in G$. Show that $\langle g \rangle$ is a subgroup of G.
2) For any subset S of G, the ***centralizer*** of S in G is
$$C_G(S) = \{ g \in G \mid g \text{ commutes with every element of } S \}.$$
   Show that $C_G(S)$ is a subgroup of G.
3) For any subgroup H of G, the ***normalizer*** of H in G is
$$N_G(H) = \{ g \in G \mid gH = Hg \},$$
   where $gH = \{ gh \mid h \in H \}$ and $Hg = \{ hg \mid h \in H \}$. Show that $N_G(H)$ is a subgroup of G.
4) The ***centre*** of G is
$$Z(G) = C_G(G) = \{ z \in G \mid zg = gz \text{ for all } g \in G \}.$$
   Show that this is an abelian, normal subgroup of G.

(1.3.5) **Exercises.**

1) Show that every cyclic group is abelian.
2) For $g \in g$, show that $|g| = |\langle g \rangle|$. More precisely, show that $\langle g \rangle = \{ g^k \mid 0 \leq k < |g| \}$, and that all of these elements are distinct.
   [*Hint:* Item 1(1b).]

(1.3.6) **Exercise.** Suppose H and K are subgroups of G, and let $HK = \{\, hk \mid h \in H,\ k \in K \,\}$. Show that HK is a subgroup of G if and only if $HK = KH$.

(1.3.7) **Definitions.**

1) For any subset S of G, we use $\langle S \rangle$ to denote the (unique) smallest subgroup of G that contains S. This is called the subgroup of G ***generated by*** S. The elements of $\langle S \rangle$ are precisely the products of the form $s_1^{k_1} s_2^{k_2} \cdots s_r^{k_r}$, where $r \in \mathbb{N}^+$ and each $s_i \in S$ and $k_i \in \mathbb{Z}$ (see Exercise 1.3.8).
2) We say that a subset S of G is a ***generating set*** for G (or that S ***generates*** G) if $\langle S \rangle = G$.

(1.3.8) **Exercise.** Let S be a nonempty subset of G, and let H be the set of all elements of G that can be written as a product of the form $s_1^{k_1} s_2^{k_2} \cdots s_r^{k_r}$, where $r \in \mathbb{N}^+$ and each $s_i \in S$ and $k_i \in \mathbb{Z}$. Show that H is the unique smallest subgroup of G that contains S. More precisely, show:

1) H is a subgroup of G that contains S, and we have $H \subseteq H'$, for every subgroup H' of G that contains S. (This is what it means to say that H is the smallest subgroup of G that contains S: it is contained in all of the others.)
2) No other subgroup of G satisfies the conditions in (1).

(1.3.9) **Definitions** (conjugates and normal subgroups)**.**

1) For any $g, h \in G$, we let
$$^g h = ghg^{-1}.$$
This is called the ***conjugate*** of H by g. Note that
$$^g h = h \quad \Leftrightarrow \quad gh = hg \quad \Leftrightarrow \quad h \text{ commutes with } g.$$
2) For any subgroup H of G and $g \in G$, we let
$$^g H = \{\, ^g h \mid h \in H \,\}.$$
This is called the ***conjugate*** of H by g. It is a subgroup of G (see Exercise 1.3.12(2)), and we have $|^g H| = |H|$ (see Exercise 1.4.6(2)).
3) We say that a subgroup K is ***conjugate*** to H if $K = {}^g H$, for some $g \in G$.
4) We say that g ***normalizes*** H if $^g H = H$.
5) A subgroup N of G is ***normal*** if every element of G normalizes N. When this is the case, we write $N \trianglelefteq G$.

(1.3.10) **Warning.** When g normalizes H, we know that $^g h$ is some element of H, for every $h \in H$. However, there is no reason to expect $^g h$ to be equal to h. Usually, conjugation by g will move the elements of H around to different places. The following is an example.

(1.3.11) **Example.** The group of rotations is a normal subgroup of D_{2n}. Any two rotations commute, but it is not difficult to see that if f is a reflection, then $^f r_\theta = r_{-\theta}$.

(1.3.12) **Exercises.**

1) Show that every subgroup of an abelian group is normal: if G is abelian and H is a subgroup of G, then $H \trianglelefteq G$.
2) Show that if H is a subgroup of G, and $g \in G$, then $^g H$ is a subgroup of G.
3) Let H be a subgroup of G. Show
$$H \trianglelefteq G \quad \Leftrightarrow \quad gH = Hg \text{ for all } g \in G \quad \Leftrightarrow \quad {}^g H \subseteq H \text{ for all } g \in G.$$
4) Show that if N_1 and N_2 are normal subgroups of G, then $N_1 \cap N_2$ is a normal subgroup of G.
5) Suppose H and K are subgroups of G. Show that if every element of H normalizes K (or vice-versa), then HK is a subgroup of G.

(1.3.13) *Remark.* Much as in Remark 1.3.3, Exercise 1.3.12(4) implies that the intersection of any finite collection of normal subgroups of G is a normal subgroup of G. But it is not difficult to show directly that the intersection of any collection of normal subgroups is also a normal subgroup, even if the collection is infinite.

(1.3.14) **Definitions** (cosets)**.** Let H be a subgroup of G, and let $g \in G$.

1) The set gH is called a ***left coset*** of H. The collection of all left cosets is denoted G/H, and it is a partition of G (see Example 1.3.17). (Recall that a ***partition*** of G is a collection of nonempty subsets of G whose union is all of G.)
2) The number of left cosets is the ***index*** of H, and is denoted $|G : H|$.

(1.3.15) **Exercises.** Let H be a subgroup of G.

1) For all $g, g_1, g_2 \in G$, show:
   (a) $g \in gH$, (b) $gH = H \Leftrightarrow g \in H$, (c) $g_1H = g_2H \Leftrightarrow g_1^{-1}g_2 \in H$.
2) Show that all cosets of H have the same cardinality: for all $g_1, g_2 \in G$, we have
$$|g_1H| = |g_2H| = |Hg_2| = |Hg_1| = |H|.$$
3) Prove Lagrange's Theorem: If G is finite, and H is a subgroup of G, then $|G : H| = |G|/|H|$. Therefore, the order of every subgroup of G is a divisor of the order of G.
4) Prove that every group of prime order is cyclic.
   [*Hint:* Lagrange's Theorem.]
5) Prove that if $|G|$ is finite, then $g^{|G|} = 1$ for all $g \in G$.
   [*Hint:* Lagrange's Theorem implies $|g| \mid |G|$.]
6) Show that every subgroup of index 2 is normal: if H is a subgroup of G, and $|G : H| = 2$, then $H \trianglelefteq G$.

(1.3.16) *Remark.* The converse of Lagrange's Theorem is not true: there are examples where G has no subgroup of order n, even though n is a divisor of $|G|$.

(1.3.17) **Example.** For any subgroup H of G, we show that the left cosets of H form a partition of G. In other words, we show that every element of G is in a unique left coset. To see this, first note that Exercise 1.3.15(1) tells us $g \in gH$. This establishes that each element of g is in a left coset, so all that remains is the uniqueness. Suppose $g \in g_1H$ and $g \in g_2H$. This means that, for each i, we may write $g = g_ih_i$, with $h_i \in H$. Therefore $g_i = gh_i^{-1}$. Also note that $h_i^{-1} \in H$ (since $h_i \in H$ and the subgroup H must be closed under inverses), so we see from Exercise 1.3.15(1) that $h_i^{-1}H = H$. Therefore
$$g_iH = (gh_i^{-1})H = g(h_i^{-1}H) = gH.$$
So $g_1H = gH = g_2H$. So any two left cosets of H that contain g are equal. This establishes the desired uniqueness of the left coset that contains g.

(1.3.18) *Remark.* The set Hg is a ***right coset*** of H. The right cosets also form a partition of G, and it is not difficult to see that the number of right cosets is equal to the number of left cosets. Some textbooks develop the theory by using right cosets instead of left cosets; which side to use is purely a matter of choice. For more advanced topics, it is often convenient to have both right cosets and left cosets available, but left cosets will suffice for our purposes this semester.

(1.3.19) **Definition.** If N is a normal subgroup of G, then G/N is a group under the operation defined by $(g_1N)(g_2N) = g_1g_2N$ (see Exercise 1.3.20(2)). This is called the ***quotient*** of G by N.

(1.3.20) **Exercises.**

1) Prove that quotients of abelian groups are abelian: if G is abelian and N is a subgroup of G, then G/N is abelian.

2) Let N be a normal subgroup of G.
   (a) Show that the binary operation on G/N is well-defined. That is, show that the product $(g_1N)(g_2N)$ depends only on the cosets g_1N and g_2N, and not on the particular representatives g_1 and g_2.
   More precisely, show that if $g_1N = g_1'N$ and $g_2N = g_2'N$, then $g_1g_2N = g_1'g_2N$.
   [*Hint:* You definitely need to use the fact that N is normal!]
   (b) Show that the binary operation on G/N is associative, and has an identity element and inverses.

§1.4. Homomorphisms and isomorphisms

(1.4.1) **Definition.** Assume G and H are groups. A function $\varphi\colon G \to H$ is a ***homomorphism*** if it respects the group operation. By this, we mean that $\varphi(g_1g_2) = \varphi(g_1)\,\varphi(g_2)$ for all $g_1, g_2 \in G$. More precisely, if the operations on G and H are $*$ and $\bullet$, respectively, then $\varphi(g_1 * g_2) = \varphi(g_1) \bullet \varphi(g_2)$.

(1.4.2) **Example.** The set $\mathbb{R}$ of real numbers is a group under addition ($+$), and the set $\mathbb{R}^+$ of positive real numbers is a group under multiplication ($\cdot$). The exponential map e^x is a homomorphism from $\mathbb{R}$ to $\mathbb{R}^\times$, because it turns addition into multiplication: $e^{x+y} = e^x \cdot e^y$.

(1.4.3) **Exercises.** Assume φ is a homomorphism from G to H.
1) By definition, we know that φ respects multiplication.
   (a) Show that φ also respects the identity element and inverses. This means that
   $$\varphi(1_G) = 1_H, \quad \text{and} \quad \varphi(g^{-1}) = \varphi(g)^{-1} \text{ for all } g \in G.$$
   (b) Show that φ also respects powers: $\varphi(g^k) = \varphi(g)^k$, for all $g \in G$ and $k \in \mathbb{Z}$.
2) Show that if K is a subgroup of G, then $\varphi(K)$ is a subgroup of H.
   (*Warning:* $\varphi(K)$ might <u>not</u> be a <u>normal</u> subgroup of H, even if K is a normal subgroup of G.)
3) We let
   $$\ker\varphi = \varphi^{-1}(1_H) = \{\, g \in G \mid \varphi(g) = 1_H \,\}.$$
   This is called the ***kernel*** of φ. Show that $\ker\varphi$ is a normal subgroup of G.
4) Conversely, suppose N is any normal subgroup of G. Show that the function $\psi(x) = xN$ is a homomorphism from G to G/N whose kernel is N.
5) Show that φ is one-to-one if and only if the kernel of φ is trivial.

(1.4.4) **Definition.**
1) A bijective homomorphism is called an ***isomorphism***.
2) We say that groups G and H are ***isomorphic*** (and write $G \cong H$) if there is an isomorphism from G to H. This is an equivalence relation (see Exercise 1.4.7(4)).

Isomorphisms preserve all properties that can be expressed in group-theoretic terms. For example:

(1.4.5) **Exercises.** Assume $\varphi\colon G \xrightarrow{\cong} H$, and let $g, g' \in G$. Then:
1) $|G| = |H|$.
2) G is abelian if and only if H is abelian.
3) g commutes with g' if and only if $\varphi(g)$ commutes with $\varphi(g')$.
4) $|\varphi(g)| = |g|$.
5) G is cyclic if and only if H is cyclic.
⋮ etc.

(1.4.6) **Exercises.**
1) Show that every cyclic group of order n is isomorphic to $\mathbb{Z}_n$.

2) Assume H is a subgroup of G, and let $g \in G$. Show $H \cong {}^gH$. (By Exercise 1.4.5(1), this implies $|H| = |{}^gH|$.)
[*Hint:* Define $\varphi\colon H \to {}^gH$ by $\varphi(h) = {}^gh$.]

(1.4.7) **Exercises.**

1) Show that the composition of homomorphisms is a homomorphism. This means that if φ is a homomorphism from G to H, and σ is a homomorphism from H to K (where G, H, and K are groups), then the composition $\sigma \circ \varphi$ is a homomorphism from G to K.
2) Show that the composition of isomorphisms is an isomorphism. This means that if φ is an isomorphism from G to H, and σ is an isomorphism from H to K (where G, H, and K are groups), then the composition $\sigma \circ \varphi$ is an isomorphism from G to K.
[*Hint:* Use previous exercises.]
3) Show that the inverse of an isomorphism is an isomorphism: if φ is an isomorphism from G to H, then φ has an inverse (which is a function from H to G), and φ^{-1} is an isomorphism from H to G.
4) Show that isomorphism is an equivalence relation on the collection of all groups. In other words, show that isomorphism is:
   - reflexive: $G \cong G$,
   - symmetric: $G \cong H \Rightarrow H \cong G$, and
   - transitive: $G \cong H \cong K \Rightarrow G \cong K$.

[*Hint:* Symmetry and transitivity follow from (3) and (2).]

(1.4.8) **Definition.** An isomorphism from G to itself is called an ***automorphism*** of G. The set of all automorphisms of G is denoted $\mathrm{Aut}(G)$. It is a group under composition (see Exercise 1.4.9(2)).

(1.4.9) **Exercises.**

1) For $g \in G$, define $\varphi_g\colon G \to G$ by $\varphi_g(x) = {}^gx$.
   (a) Show that φ_g is an automorphism of G. It is called the ***conjugation by g*** (or the ***inner automorphism*** corresponding to g).
   (b) Show that the map $g \mapsto \varphi_g$ is a homomorphism from G to $\mathrm{Aut}\, G$.
2) Show that $\mathrm{Aut}(G)$ is a group under composition.
3) For $n, k \in \mathbb{Z}^+$, define $\varphi_k\colon \mathbb{Z}_n \to \mathbb{Z}_n$ by $\varphi_k(x) = kx$.
   (a) Show that φ_k is a homomorphism.
   (b) Show that φ_k is an automorphism of $\mathbb{Z}_n$ if and only if $\gcd(k, n) = 1$.
   (c) For each $\varphi \in \mathrm{Aut}(\mathbb{Z}_n)$, show there is some $k \in \mathbb{Z}^+$, such that $\varphi = \varphi_k$.
   (d) Show that $\varphi_k = \varphi_\ell$ if and only if $k \equiv \ell \pmod{n}$.

(1.4.10) **Proposition** (1st, 2nd, and 3rd Isomorphism Theorems)**.**

1) *If φ is a homomorphism from G to H, then $G/\ker\varphi \cong \varphi(G)$.*
   More precisely, an isomorphism $\overline{\varphi}\colon G/\ker\varphi \to \varphi(G)$ can be obtained by defining
   $$\overline{\varphi}(g\ker\varphi) = \varphi(g).$$
2) *Suppose $N \trianglelefteq G$ and H is a subgroup of G. Then*
   (a) *HN is a subgroup of G (where $HN = \{\, hn \mid h \in H,\ n \in N \,\}$),*
   (b) *$H \cap N$ is a normal subgroup of H, and*
   (c) *$HN/N \cong H/(H \cap N)$.*
3) *Suppose H and N are normal subgroups of G, with $N \subseteq H$. Then H/N is a normal subgroup of G/N, and $(G/N)/(H/N) \cong G/H$.*

Sketch of proof. (1) For $g \in G$, we have
$$\varphi(g\ker\varphi) = \varphi(g)\cdot\varphi(\ker\varphi) = \varphi(g)\cdot\{1\} = \{\varphi(g)\},$$
so $\overline{\varphi}$ is well defined. Also, if $\overline{\varphi}(g\ker\varphi) = 1$, then $\varphi(g) = \overline{\varphi}(g\ker\varphi) = 1$, which means $g \in \ker\varphi$, so $g\ker\varphi = \ker\varphi$ is the trivial coset; therefore $\ker\overline{\varphi}$ is trivial, so $\overline{\varphi}$ is one-to-one. Finally, for $h \in \varphi(G)$, there exists $g \in G$, such that $\varphi(g) = h$, so $\overline{\varphi}(g\ker\varphi) = \varphi(g) = h$; therefore $\overline{\varphi}$ is onto.

(2a) It suffices to show that HN is closed under multiplication and inverses. Since $N \trianglelefteq G$, we have $hN = Nh$ for all $h \in H$; therefore $HN = NH$. Hence,

$$(HN)(HN) = H(NH)N = H(HN)N = (HH)(NN) = HN,$$

so HN is closed under multiplication. Also,

$$(HN)^{-1} = N^{-1}H^{-1} = NH = HN,$$

so HN is closed under inverses.

(2b) For $h \in H$, we have $h(H \cap N)h^{-1} = (hHh^{-1}) \cap (hNh^{-1}) = H \cap N$, so $H \cap N \trianglelefteq H$.

(2c) Let $\varphi\colon G \to G/N$ be the natural homomorphism with kernel N, and let φ' be the restriction of φ to H. Then $\ker \varphi' = H \cap \ker \varphi = H \cap N$, so applying (1) to the homomorphism φ' yields

$$HN/N = \varphi'(H) \cong H/\ker \varphi' = H/(H \cap N).$$

(3) Let $\varphi\colon G \to G/H$ be the natural homomorphism. The proof of (1) implies that there is a well-defined surjective homomorphism $\overline{\varphi}\colon G/N \to G/H$, defined by $\overline{\varphi}(gN) = \varphi(g)$, and that $\ker \overline{\varphi} = H/N$. Then applying (1) to the homomorphism $\overline{\varphi}$ yields

$$(G/N)/(H/N) = (G/N)/\ker \overline{\varphi} \cong \overline{\varphi}(G/N) = G/H. \qquad \square$$

(1.4.11) *Remark.* When you need to show that a quotient group G/N is isomorphic to some group H, you should almost never try to directly construct an isomorphism from G/N to H. Instead, Proposition 1.4.10(1) says that it suffices to find a homomorphism from G onto H whose kernel is N. This is usually much easier.

(1.4.12) **Proposition** (Correspondence Theorem)**.** *Suppose N is a normal subgroup of G. Then there is a one-to-one correspondence between the subgroups of G that contain N and the subgroups of G/N. Namely, the subgroup of G/N corresponding to a subgroup H of G is H/N.*

Furthermore, this remains a one-to-one correspondence when restricted to the normal subgroups of G and G/N.

Sketch of proof. Let

- $\mathcal{H}$ be the collection of all subgroups of G that contain N,
- $\overline{\mathcal{H}}$ be the collection of all subgroups of G/N, and
- $\varphi\colon G \to G/N$ be the natural homomorphism with kernel N.

Define:

- $\overline{\varphi}\colon \mathcal{H} \to \overline{\mathcal{H}}$ by $\overline{\varphi}(H) = \varphi(H)$ $(= H/N)$, and
- $\widehat{\varphi}\colon \overline{\mathcal{H}} \to \mathcal{H}$ by $\widehat{\varphi}(K) = \varphi^{-1}(K)$.

It is straightforward to verify that $\overline{\varphi}$ and $\widehat{\varphi}$ are inverses of each other, so $\overline{\varphi}$ is a bijection.

For $K \in \overline{\mathcal{H}}$, it is also straightforward to verify that $K \trianglelefteq G/N \Leftrightarrow \widehat{\varphi}(K) \trianglelefteq G$. $\square$

Chapter 2

Group Actions

§2.1. Definition and basic facts

(2.1.1) **Definition.** A (left) ***action*** of a group G on a set X is a function $\alpha\colon G \times X \to X$ that satisfies the following two axioms. (We often write $g * x$ or gx or ${}^g x$ as shorthand for $\alpha(g, x)$.)

1) $g(hx) = (gh)x$ for all $g, h \in G$ and $x \in X$, and
2) $1x = x$ for all $x \in X$ (where 1 is the identity element of G).

Here is another way to think about group actions:

(2.1.2) **Lemma.** *Suppose G acts on X, and S_X is the group of all permutations of X. For each $g \in G$, define $\varphi_g\colon X \to X$ by $\varphi_g(x) = gx$. Then*

1) *$\varphi_g \in S_X$, and*
2) *$\varphi\colon G \to S_X$ is a homomorphism.*

Idea of proof. (1) Verify that $\varphi_{g^{-1}}$ is the inverse of φ_g:

$$\varphi_{g^{-1}}(\varphi_g(x)) = \varphi_{g^{-1}}(gx) = g^{-1}(gx) = (g^{-1}g)x = 1x = x.$$

Any function with an inverse is a bijection, so this implies that φ_g is a bijection.

(2) Verify that $\varphi_{gh} = \varphi_g \varphi_h$: for $x \in X$, we have

$$\varphi_{gh}(x) = (gh)x = g(hx) = \varphi_g(hx) = \varphi_g(\varphi_h(x)) = (\varphi_g \varphi_h)(x). \qquad \square$$

(2.1.3) **Exercise.** Conversely, verify that every homomorphism $\varphi\colon G \to S_X$ yields an action of G on X, by defining $gx = \varphi_g(x)$.

(2.1.4) **Definition.** Recall that every homomorphism φ has a ***kernel***:

$$\ker \varphi = \{\, g \in G \mid \varphi(g) = 1 \,\}.$$

The ***kernel*** of a group action is the kernel of the corresponding homomorphism:

$$\{\, g \in G \mid \varphi_g \text{ is the trivial permutation} \,\} = \{\, g \in G \mid \forall x \in X,\ gx = x \,\}.$$

(2.1.5) **Example** (regular action)**.** G acts on itself by multiplication on the left:

for $X = G$, we may let $g * x = gx$ (multiplication in the group).

You probably saw this action in your undergraduate abstract algebra class, because it is used in the proof of the following fundamental result:

(2.1.6) **Basic fact** (Cayley's Theorem)**.** Every finite group is isomorphic to a subgroup of S_n, for some n. More precisely, we may take $n = |G|$.

You should verify for yourself that this (and the following examples) satisfy the conditions to be an action. (For example, the above example uses (1) the fact that groups are associative, and (2) the definition of the identity element.)

(2.1.7) **Examples.** Every group has the following actions (in addition to the regular action that was described in Example 2.1.5):

1) G acts ***trivially*** on any set X, by letting $gx = x$ for all $g \in G$ and $x \in X$.
2) G acts on itself by conjugation, letting ${}^g x = gxg^{-1}$. (On the right-hand side, the multiplication is the group operation of G.)
3) For any subgroup H of G, recall that G/H is the set of left cosets of H in G:
$$G/H = \{\, gH \mid g \in G \}.$$
G acts on this set by left multiplication: $g * xH = gxH$.

(2.1.8) **Examples.**

1) As in Lemma 2.1.2 above, let S_X be the group of all permutations of X. Then S_X acts on X in a natural way. (The identity map is a homomorphism from $G = S_X$ to S_X.) It also acts on the power set $\mathcal{P}(X)$ (the set of all subsets of X), by $\sigma A = \{\, \sigma(a) \mid a \in A \,\}$ for $\sigma \in S_X$ and $A \subseteq X$.
2) More generally, if G acts on X, then G also acts on $\mathcal{P}(X)$, by $gA = \{\, ga \mid a \in A \,\}$ for $A \subseteq X$.
3) Let H be a group. If G acts on X, and we have a homomorphism $\varphi\colon H \to G$, then an action of H on X can be defined by letting $hx = \varphi(h)\, x$.
4) Any action of G on X can be restricted to any subgroup of G. More precisely, if G acts on X, and H is a subgroup of G, then an action of H on X is obtained by defining $h * x = hx$.

(2.1.9) *Other conventions.* Some authors use right actions, instead of left actions, so $\alpha\colon X \times G \to X$, and the axioms are: $(xg)h = x(gh)$ and $x1 = x$. To compensate for this, the action by conjugation is defined by $x^g = g^{-1}xg$.

§2.2. Orbits and stabilizers

(2.2.1) **Assumption.** In this section, we assume that G acts on X.

(2.2.2) **Definitions.** Let $x \in X$.

1) The ***orbit*** of x under G is $Gx = \{\, gx \mid g \in G \,\}$.
2) The ***stabilizer*** of x in G is $\mathrm{Stab}_G(x) = \{\, g \in G \mid gx = x \,\}$.
3) If $Gx = X$ for some (or, equivalently, all) $x \in X$, then we say that the action is ***transitive***.

(2.2.3) **Exercises.**

1) Show that G is transitive on X if and only if for all $x, y \in X$, there exists $g \in G$, such that $gx = y$.
2) Show that the orbits form a ***partition*** of X. That is, every element of X is in a unique orbit (and no orbit is the empty set). In other words, show that
   (a) the union of all of the orbits is X,
   (b) any two different orbits are disjoint, and
   (c) no orbit is the empty set.
3) Show $\mathrm{Stab}_G(x)$ is a subgroup of G, for all $x \in X$.
4) For $g \in G$ and $x \in X$, show $\mathrm{Stab}_G(gx) = g\, \mathrm{Stab}_G(x)\, g^{-1}$. (This implies that if x and y are in the same orbit, then $\mathrm{Stab}_G(x)$ is conjugate to $\mathrm{Stab}_G(y)$, so the two stabilizers have the same order.)
5) Suppose $N \trianglelefteq G$, and $N \subseteq \mathrm{Stab}_G(x)$, for some $x \in X$. Show that if G is transitive on X, then N is contained in the kernel of the action.

6) Suppose
   (a) $x \in X$, $g, h \in G$, and
   (b) K is a subgroup of G that contains $\mathrm{Stab}_G(x)$.
   Let
   $$Y = Kx = \{ kx \mid k \in K \}, \quad gY = \{ gy \mid y \in Y \}, \quad hY = \{ hy \mid y \in Y \}.$$
   Show that if gY is not disjoint from hY, then $gY = hY$.

(2.2.4) **Examples.**

1) For the regular action of G (by left-multiplication on itself):
   - The action is transitive. (For $x, y \in G$, if we let $g = yx^{-1}$, then $y = gx$.)
   - $\mathrm{Stab}_G(x)$ is trivial. (If $gx = x$, then $g = (gx)x^{-1} = xx^{-1} = 1$.)

   In this case, the orbit is large (all of G) and the stabilizer is very small (trivial).
2) At the other extreme, for the trivial action of G on X:
   - Each orbit is only a single point. (If $gx = y$, then $x = y$.)
   - $\mathrm{Stab}_G(x) = G$. (We have $gx = x$ for all g.)

   In this case, every orbit is small (a single point), and the stabilizers are large (all of G).
3) Let $A \subseteq X$, and let $k = \#A$. For the action of S_X on $\mathcal{P}(X)$:
   - We have $S_X A = \{ B \in \mathcal{P}(X) \mid \#B = k \}$. So the number of orbits is exactly $1 + \#X$ (because k can be any number from 0 to $\#X$).
   - $\mathrm{Stab}_{S_X}(A) \cong S_A \times S_{X \smallsetminus A}$.

(2.2.5) **Exercise.** For any subgroup H of G, show the action of G on G/H (by left multiplication) is transitive, and that $\mathrm{Stab}_G(gH) = gHg^{-1}$, for all $g \in G$.

The following important result shows that points with large stabilizers always have small orbits, and points with small stabilizers have large orbits. More precisely, the cardinality of the orbit is the index of the stabilizer:

(2.2.6) **Theorem** (Orbit-Stabilizer Theorem)**.** *Assume G acts on X, and $x \in X$. Then*
$$|Gx| = |G : \mathrm{Stab}_G(x)|.$$

Proof. For convenience, let $H = \mathrm{Stab}_G(x)$, so $|G : \mathrm{Stab}_G(x)| = |G/H|$. Define $f\colon G/H \to Gx$ by $f(gH) = gx$. Then f is:

- onto: Given $y \in Gx$, there exists $g \in G$, such that $y = gx = f(gH)$.
- one-to-one: If $f(g_1H) = f(g_2H)$, then $g_1x = g_2x$, so (multiplying both sides on the left by g_2^{-1}, we have $g_2^{-1}g_1x = x$. This means $g_2^{-1}g_1 \in \mathrm{Stab}_G(x) = H$, so $g_1H = g_2H$.
- well-defined: If $g_1H = g_2H$, then $g_2^{-1}g_1 \in H = \mathrm{Stab}_G(x)$, so
  $$f(g_1H) = g_1x = g_2(g_2^{-1}g_1)x = g_2x = f(g_2H).$$

Thus, f is a bijection from G/H to Gx, so the two sets have the same cardinality. □

Here is another way of saying the same thing:

(2.2.7) **Corollary.** $|\mathrm{Stab}_G(x)| \cdot |Gx| = |G|$.

(2.2.8) *Remark.* Lagrange's Theorem tells us $|\mathrm{Stab}_G(x)|$ is a divisor of G (if G is finite). The corollary tells us that $|Gx|$ is also a divisor of G.

(2.2.9) **Example.** The group Rot(cube) of all rotations of a cube acts transitively on the six faces of the cube. (If you pick up a cube, you can set it down into the exactly the same space it was in before, with any face you choose on the bottom.) And there are exactly four rotations of the cube that keep a given face fixed (setwise). So the Orbit-Stabilizer Theorem implies that the order of Rot(cube) is $6 \times 4 = 24$.

(2.2.10) **Exercises.**

1) Show that Rot(cube) is isomorphic to the symmetric group S_4 on four symbols.
   [*Hint:* Rot(cube) permutes the four diagonals of the cube.]

2) Show that if H and K are subgroups of G, and G is finite, then
$$|HK| = |H|\,|K|/|H \cap K|.$$
[*Hint:* Restrict the action of G on G/K to H, and apply the Orbit-Stabilizer Theorem. We have $\mathrm{Stab}_H(K) = H \cap K$, and the cardinality of the H-orbit of K is $|HK|/|K|$.]

3) Assume
(a) p is a prime number that divides $|G|$, but does not divide $|X|$, and
(b) p^r is the largest power of p that divides $|G|$.
Show there exists $x \in X$, such that $|\mathrm{Stab}_G(x)|$ is divisible by p^r.

(2.2.11) **Application.** Let us apply the Orbit-Stabilizer Theorem to the action of G on itself by conjugation: ${}^g x = gxg^{-1}$. Each orbit of this action is called a ***conjugacy class*** in G. Since the orbits (or conjugacy classes) partition G, we have
$$|G| = \sum_{\text{conjugacy classes } C} |C| = \begin{pmatrix}\text{number of conjugacy}\\ \text{classes of cardinality 1}\end{pmatrix} + \sum_{\substack{\text{conjugacy classes } C\\ \text{with } |C|>1}} |C|.$$
Note that:

- The conjugacy class of x has cardinality 1 if and only if x commutes with every element of G, which means $x \in Z(G)$ (the centre of G). So the number of conjugacy classes of cardinality 1 is $|Z(G)|$.
- The Orbit-Stabilizer Theorem tells us $|C| = |G : \mathrm{Stab}_G(x)|$ for $x \in C$.
- We have $g \in \mathrm{Stab}_G(x)$ if and only if g commutes with x, which means $g \in C_G(x)$ (the centralizer of x).

Therefore, if we choose an element x_i from each conjugacy class of cardinality > 1, then
$$|G| = |Z(G)| + \sum_i |G : C_G(x_i)|. \tag{2.2.12}$$
This is the ***class equation***. It is an important tool in the theory of finite groups.

We can also now complete some unfinished business from Section 1.2:

Proof of Burnside's Counting Lemma (1.2.3). This proof employs a standard technique: count the elements of a set in two different ways. Let
$$\mathcal{F} = \{\,(g,x) \in G \times X \mid gx = x\,\}.$$
For each $g \in G$, let $f(g)$ be the number of fixed points of g, so
$$f(g) = \#\{\,x \in X \mid (g,x) \in \mathcal{F}\,\}.$$
Therefore
$$\#\mathcal{F} = \sum_{g \in G} f(g).$$
Also, for any $x \in X$, we have $\mathrm{Stab}_G(x) = \{\,g \in G \mid (g,x) \in \mathcal{F}\,\}$, so
$$\#\mathcal{F} = \sum_{x \in X} |\mathrm{Stab}_G(x)|.$$
Therefore
$$\sum_{g \in G} f(g) = \sum_{x \in X} |\mathrm{Stab}_G(x)|. \tag{2.2.13}$$
Recall that when x and y are in the same orbit, we have $|\mathrm{Stab}_G(x)| = |\mathrm{Stab}_G(y)|$ (see Exercise 2.2.3(4)). Hence, for each $x \in X$, we have
$$\sum_{y \in Gx} |\mathrm{Stab}_G(y)| = \sum_{y \in Gx} |\mathrm{Stab}_G(x)| = |Gx| \cdot |\mathrm{Stab}_G(x)| = |G| \tag{2.2.14}$$
(by the Orbit-Stabilizer Theorem). We can break up the right-hand-side of (2.2.13) into sums over orbits, and (2.2.14) shows that the sum over each orbit is $|G|$. So we conclude that the right-hand side of (2.2.13) is equal to $|G|$ times the number of orbits in X. Dividing by $|G|$, we find that $(1/|G|)\sum_{g \in G} f(g)$ equals the number of orbits, as desired. □

§2.3. Sylow Theorems

(2.3.1) **Definitions.** Suppose G is finite, p is a prime number, and P is a subgroup of G.

1) G is a ***p*-group** if $|G|$ is a power of p. (That is $|G| = p^k$, for some $k \in \mathbb{N}$.)
2) P is a ***p*-subgroup** of G if $|P|$ is a power of p. (In other words, a p-subgroup is a subgroup that also happens to be a p-group.)
3) P is a ***Sylow p*-subgroup** of G if $|P|$ is the largest power of p that divides $|G|$.
4) $\mathbf{Syl}_p(G)$ is the set of all Sylow p-subgroups of G.
5) P is a ***Sylow* subgroup** of G if it is a Sylow p-subgroup of G for some prime p.

(2.3.2) **Example.** Suppose $|G| = 600 = 2^3 \cdot 3 \cdot 5^2$. Then the Sylow p-subgroups of G are the subgroups of order:

$$2^3 \text{ if } p = 2, \quad 3 \text{ if } p = 3, \quad 5^2 \text{ if } p = 5, \quad 1 \text{ if } p > 5.$$

So the Sylow subgroups of G are the subgroups of order 1, 3, 8, or 25.

(2.3.3) *Remark.* By Lagrange's Theorem, we know that if P is any subgroup of G, then $|P|$ is a divisor of $|G|$. Thus, to say that P is a Sylow p-subgroup means that it is a p-subgroup of the largest possible order that is compatible with Lagrange's Theorem — there could not possibly be a p-subgroup of larger order.

In this section, we establish the following fundamental result that is often stated, but usually not proved, in the first semester undergraduate abstract algebra.

(2.3.4) **Theorem** (Sylow Theorems)**.** *Let G be a finite group and let p be a prime number. Then:*

1) (existence) *G has a Sylow p-subgroup.*
2) (conjugacy) *Any two Sylow p-subgroups of G are conjugate.*
3) (development) *Every p-subgroup of G is contained in a Sylow p-subgroup of G.*
4) *The number of Sylow p-subgroups of G is a divisor of $|G|$, and is congruent to 1, modulo p.*

Our proof of the existence of Sylow p-subgroups will use the following fact from elementary number theory.

(2.3.5) **Lemma.** *Assume p is prime, $r \in \mathbb{N}$, and $m \in \mathbb{Z}^+$, such that $p \nmid m$. Then $p \nmid \binom{p^r m}{p^r}$, where the notation $\binom{n}{k}$ denotes the* ***binomial coefficient*** $\frac{n!}{k!\,(n-k)!}$.

We sketch two of the numerous different proofs of this fact.

Proof 1. By induction on r, it suffices to show $\binom{pn}{pk} \equiv \binom{n}{k} \pmod{p}$ for $n, k \in \mathbb{N}$. Since p is prime, it is well known (and easy to prove from the fact that $p \mid \binom{p}{k}$ for $1 \le k \le p-1$) that $(a+b)^p \equiv a^p + b^p \pmod{p}$. So

$$(1+x)^{pn} \equiv (1+x^p)^n \pmod{p}.$$

By the binomial theorem, the coefficient of x^{pk} on the left-hand side is $\binom{pn}{pk}$, but the coefficient on the right-hand side is $\binom{n}{k}$. So these two binomial coefficients must be congruent, modulo p. □

Proof 2. We have

$$\binom{p^r m}{p^r} = \frac{(p^r m)!}{(p^r)!\,(p^r m - p^r)!} = \prod_{i=0}^{p^r-1} \frac{p^r m - i}{p^r - i}.$$

In each of the factors in this product, elementary number theory reveals that the largest power of p that divides the numerator is exactly the same as the largest power of p that divides the denominator. That is, all occurrences of p cancel, so no factor in this product is divisible by p. Hence, the product is not divisible by p. □

(2.3.6) *Remark.* Actually, something much stronger than Lemma 2.3.5 is true. Namely, suppose we write n and r in base p (where p is prime):

$$n = n_r p^r + n_{r-1}p^{r-1} + \cdots + n_1 p + n_0 \text{ and } k = k_r p^r + k_{r-1}p^{r-1} + \cdots + k_1 p + k_0,$$

with n_i and k_i in $\{0, 1, \ldots, p-1\}$. Then

$$\binom{n}{k} \equiv \binom{n_r}{k_r}\binom{n_{r-1}}{k_{r-1}} \cdots \binom{n_0}{k_0} \pmod{p}.$$

This can be established by the method of Proof 1.

Proof of Sylow's Existence Theorem. Let p^r be the largest power of p that divides $|G|$. We wish to show that G has a subgroup of order p^r.

To accomplish this, let

$$X = \{ A \subseteq G \mid |A| = p^r \}$$

be the collection of all subsets (not just subgroups) of G that have cardinality p^r. Then G acts on X by left multiplication: $g \cdot A = \{ ga \mid a \in A \}$. Since $|X| = \binom{|G|}{p^r}$, and $|G|/p^r$ is not divisible by p, Lemma 2.3.5 tells us that $|X|$ is not divisible by p, so there must an orbit $G \cdot A$ whose cardinality is not divisible by p. Since

$$|\mathrm{Stab}_G(A)| = \frac{|G|}{|G \cdot A|},$$

we see that p^r is a divisor of $|\mathrm{Stab}_G(A)|$.

On the other hand, for any $a \in A$, we have $\mathrm{Stab}_G(A)a \subseteq \mathrm{Stab}_G(A)\, A = A$, so

$$|\mathrm{Stab}_G(A)| \leq |Aa^{-1}| = |A| = p^r.$$

Therefore, we must have $|\mathrm{Stab}_G(A)| = p^r$, so $\mathrm{Stab}_G(A)$ is a Sylow p-subgroup of G. □

The following lemma is a crucial tool in the remaining parts of the proof.

(2.3.7) **Definition.** Suppose G acts on X, and $x \in X$. We say x is a ***fixed point*** of the action if $gx = x$ for all $g \in G$. In other words, $|Gx| = 1$.

(2.3.8) **Lemma.** *For any action of a p-group P on a finite set X, the number of fixed points is congruent to $|X|$, modulo p.*

Proof. Write $X = F \sqcup Y$ (disjoint union), where F is the set of fixed points, and Y is the union of the orbits of cardinality > 1. The Orbit-Stabilizer Theorem tells us that the cardinality of every orbit is a divisor of $|P|$, and is therefore a power of p. Hence, the cardinality of each orbit in Y is multiple of p (since 1 is the only power of p that is not divisible by p). Since these orbits form a partition of Y, we conclude that $|Y|$ is a sum of multiples of p, and is therefore itself a multiple of p. So $|F| = |X| - |Y| \equiv |X| \pmod{p}$. □

Proof of Sylow's Conjugacy Theorem. Let $P, Q \in \mathrm{Syl}_p(G)$. We wish to show that Q is conjugate to P. To do this, we begin by repeating part of the proof of Sylow's Development Theorem.

The group G acts on G/P by multiplication on the left: $g * (hP) = ghP$ (see Example 2.1.7(3)). Restrict this to an action of Q on G/P. Since Q is a p-group and $|G/P| = |G|/|P|$ is not divisible by p (since P is a Sylow p-subgroup), we see from Lemma 2.3.8 that Q must have at least one fixed point gP in G/P. This means $Q \subseteq \mathrm{Stab}_G(gP) = {}^gP$.

However, since P and Q are Sylow p-subgroups, we know that $|P| = |Q|$. Also, since conjugate subgroups have the same order (see Exercise 1.4.6(2)), we have $|{}^gP| = |P|$. Therefore $|Q| = |P| = |{}^gP|$. Since $Q \subseteq {}^gP$ (and the groups are finite), this implies $Q = {}^gP$. Thus, we have shown that Q is a conjugate of P, as desired. □

Proof of Sylow's Development Theorem. Given any p-subgroup Q of G, we wish to show that Q is contained in some Sylow p-subgroup. By Sylow's Existence Theorem, we may fix some $P \in \mathrm{Syl}_p(G)$. The group G acts on G/P by multiplication on the left: $g * (hP) = ghP$ (see Example 2.1.7(3)). Restrict this to an action of Q on G/P. Since Q is a p-group and $|G/P| = |G|/|P|$ is not divisible by p (since P is a Sylow p-subgroup), we see from Lemma 2.3.8 that Q must have at least one fixed

point gP in G/P. This means $Q \subseteq \mathrm{Stab}_G(gP) = {}^gP$. Also, since conjugate subgroups have the same order (see Exercise 1.4.6(2)), and P is a Sylow p-subgroup, we know that ${}^gP \in \mathrm{Syl}_p(G)$. Therefore, we have shown that Q is contained in the Sylow p-subgroup ${}^gP$. □

Our proof of the final part will use the following observation.

(2.3.9) **Exercise.** Let P be a Sylow p-subgroup of G and suppose Q is a p-subgroup of $N_G(P)$ (the normalizer of P). Show $Q \subseteq P$.
[*Hint:* PQ is a p-subgroup of G that contains P.]

(2.3.10) **Exercise.** Let P and Q be Sylow p-subgroups of G. If $P \subseteq N_G(Q)$, then $P = Q$.
[*Hint:* PQ is a p-subgroup of G that contains P.]

Proof of part (4) of Sylow's Theorem. Fix some $P \in \mathrm{Syl}_p(G)$. Since conjugate subgroups have the same order, G acts on the set $\mathrm{Syl}_p(G)$ by conjugation. Furthermore, Sylow's Conjugacy Theorem tells us that this action is transitive. Therefore, the Orbit-Stabilizer Theorem tells us $|\mathrm{Syl}_p(G)| = |G : N_G(P)|$. This is a divisor of $|G|$.

Restricting this conjugation action to P yields an action of P by conjugation on $\mathrm{Syl}_p(G)$. From Exercise 2.3.9, we see that the only fixed point is P. So Lemma 2.3.8 tells us that $|\mathrm{Syl}_p(G)| \equiv 1 \pmod{p}$. □

(2.3.11) **Exercises.** Assume G is finite, p is a prime number, and $P \in \mathrm{Syl}_p(G)$.

1) Show the following are equivalent:
   (a) $P \trianglelefteq G$.
   (b) P is the unique Sylow p-subgroup of G.
2) Let n be the number of Sylow p-subgroups of G. Show G has a subgroup of index n.
3) Suppose $N \trianglelefteq G$. Show:
   (a) $P \cap N \in \mathrm{Syl}_p(N)$.
   (b) $PN/N \in \mathrm{Syl}_p(G/N)$.

(2.3.12) **Exercises.**

1) Show that if $G/Z(G)$ is cyclic, then G is abelian.
2) Let p be a prime number. Show that every group of order p^2 is abelian.
   [*Hint:* Groups of prime order are cyclic. Use Exercise 1.]
3) Assume p and q are prime numbers, with $p > q$. Show that if $p \not\equiv 1 \pmod{q}$, then every group of order pq is abelian.
   [*Hint:* Let P be a Sylow p-subgroup of G. Since 1 is the only divisor of $|G|$ that is $\equiv 1 \pmod{p}$, we know $P \trianglelefteq G$, so G acts on P by conjugation. However, $|\mathrm{Aut}(P)| = p - 1$ (see Exercise 1.4.9(3)), which is relatively prime to $|G|$, so the action must be trivial, which means $P \subseteq Z(G)$.]
4) Assume p and q are prime numbers, with $p > q$. Show that every group of order pq is isomorphic to the semidirect product $\mathbb{Z}_p \rtimes_k \mathbb{Z}_q$, for some $k \in \mathbb{Z}^+$. (See Exercise 1.1.8 for the definition of the semidirect product.)
5) Recall that A_4 is the alternating group of degree 4. (This has order 12, because it is a subgroup of index 2 in the symmetric group S_4, which has order $4! = 24$.)
   (a) How many Sylow 3-subgroups does A_4 have?
   (b) How many elements of order 3 does A_4 have?
   (c) How many Sylow 2-subgroups does A_4 have?
6) Show that if $|P| = p^n$, then P has a normal subgroup of order p^k, for $0 \leq k \leq n$.
   [*Hint:* Exercise 2.3.13(1), induction on $|P|$, and the Correspondence Theorem.]
7) Suppose Q is a p-subgroup of G, and $Q \trianglelefteq G$. Show that Q is contained in every Sylow p-subgroup of G.

(2.3.13) **Exercises.** Use Lemma 2.3.8 to prove:

1) Every nontrivial p-group has nontrivial centre. That is, if P is a p-group, and $P \neq \{1\}$, then $Z(P) \neq \{1\}$.

   [*Hint:* Let P act on itself by conjugation (${}^g x = gxg^{-1}$). Fixed points are precisely the elements of $Z(P)$, and we know that $1 \in Z(P)$.]

2) (Cauchy's Theorem) If $|G|$ is divisible by p, then G has an element of order p.

   [*Hint:* Let
   $$X = \{\, (g_1, g_2, \ldots, g_p) \in G^p \mid g_1 g_2 \cdots g_p = 1 \,\}.$$
   Then $\mathbb{Z}_p$ acts on X by cyclic permutations:
   $$k * (g_1, g_2, \ldots, g_p) = (g_{k+1}, g_{k+2}, \ldots, g_{k+p}),$$
   where the subscripts are read modulo p. Since $|X| = |G|^{p-1}$ is divisible by p, the number of fixed points must be divisible by p. One fixed point is $(1, 1, \ldots, 1)$. Any other fixed point is of the form $(g, g, \ldots, g)$, where g is an element of order p.]

3) (Fermat's Little Theorem) If p is prime and $k \in \mathbb{Z}$, then $k^p \equiv k \pmod{p}$.

   [*Hint:* Assume, for simplicity, that $k \in \mathbb{Z}^+$. Then the group $\mathbb{Z}_p$ acts by rotations on the set of all possible colourings of the vertices of the regular p-gon with k colours of paint.]

It is also instructive to have alternate proofs of the results in Exercise 2.3.13:

(2.3.14) **Exercises.**

1) Derive Exercise 2.3.13(1) from the class equation (2.2.12).

   [*Hint:* If $x \notin Z(P)$, then $|P : C_P(x)|$ is divisible by p (why?).]

2) Derive Cauchy's Theorem 2.3.13(2) from Sylow's Theorem.

   [*Hint:* Let P be a Sylow p-subgroup of G, and let g be any nontrivial element of P. Lagrange's Theorem tells us that $|g| = p^k$ for some k. Then g^{k-1} is an element of order p.]

3) Derive Fermat's Little Theorem 2.3.13(3) from Lagrange's Theorem.

   [*Hint:* Assume $k \not\equiv 0 \pmod{p}$, so $k \in \mathbb{Z}_p^\times$, the group of units of the ring $\mathbb{Z}_p$. Since $|\mathbb{Z}_p^\times| = p - 1$, Lagrange's Theorem implies $k^{p-1} \equiv 1 \pmod{p}$.]

Chapter 3

Series of subgroups

§3.1. Solvable groups and subnormal series

(3.1.1) **Definition.** A subgroup H of G is ***characteristic*** in G if $\varphi(H) = H$ for all $\varphi \in \mathrm{Aut}(G)$. We may write $H \operatorname{char} G$ to denote that H is characteristic in G.

Since conjugation by any element of G is an automorphism, we have the following simple observation:

(3.1.2) **Exercise.** Show that if $H \operatorname{char} G$, then $H \trianglelefteq G$.

The converse of Exercise 3.1.2 is usually not true, but it holds for Sylow subgroups:

(3.1.3) **Exercise.** For $P \in \mathrm{Syl}_p(G)$, show $P \trianglelefteq G \Leftrightarrow P \operatorname{char} G$.

[*Hint:* Recall that all Sylow p-subgroups are conjugate. What does that say if $P \trianglelefteq G$?]

(3.1.4) **Example.** The centre of any group is characteristic.

Proof. Let $\varphi \in \mathrm{Aut}(G)$. For $z \in Z(G)$ and $g \in G$, we have $gz = zg$. Applying φ to both sides yields $\varphi(g)\varphi(z) = \varphi(z)\varphi(g)$, which means that $\varphi(z)$ commutes with $\varphi(g)$. Now $\varphi(g)$ is an arbitrary element of G (since φ is an automorphism, and therefore onto), so this implies $\varphi(z) \in Z(G)$. Then, since z is an arbitrary element of $Z(G)$, this implies $\varphi(Z(G)) \subseteq Z(G)$.

By the same argument (with φ^{-1} in the place of φ), we have $\varphi^{-1}(Z(G)) \subseteq Z(G)$. Applying φ to both sides yields $Z(G) \subseteq \varphi(Z(G))$. Since we also know $\varphi(Z(G)) \subseteq Z(G)$ (from the preceding paragraph), we conclude that $\varphi(Z(G)) = Z(G)$. □

(3.1.5) *Remark.* Example 3.1.4 is just one instance of the general principle that any subgroup that is defined directly from G is characteristic. (The definition cannot depend on the choice of a specific element or other part of G. For example, although the centre of G is characteristic, the centralizer of a particular element of G will usually not be characteristic. Namely, $\varphi(C_G(g))$ is equal to $C_G(\varphi(g))$, the centralizer of $\varphi(g)$, which is not usually equal to the centralizer of g.)

(3.1.6) **Exercise.** For $\varphi \in \mathrm{Aut}(G)$ and $g \in G$, show $\varphi(C_G(g)) = C_G(\varphi(g))$.

(3.1.7) **Definition.**

1) For $g, h \in G$, the ***commutator*** of g and h is
$$[g,h] = ghg^{-1}h^{-1}.$$
Note that g commutes with h if and only if $[g,h] = 1$. In general, we have $gh = [g,h]\,hg$.
2) The ***commutator subgroup*** of G is
$$[G,G] = \langle [g,h] \mid g,h \in G \rangle.$$
3) More generally, for any subgroups H and K of G, we let
$$[H,K] = \langle [h,k] \mid h \in H,\ k \in K \rangle.$$

(3.1.8) *Other conventions.* Note that $[g,h] = {}^{g}h\,h^{-1}$. Authors who use right actions, instead of left actions, usually compensate by defining $[g,h]$ to be $g^{-1}h^{-1}gh = g^{-1}\,g^{h}$.

(3.1.9) **Exercise.**

1) Since $[G,G]$ has been defined from G, it should be characteristic. Show this is true: $[G,G] \operatorname{char} G$. (This implies $[G,G] \trianglelefteq G$.)
2) Show G is abelian if and only if $[G,G]$ is trivial.
3) Show $G/[G,G]$ is abelian.
4) More generally, for every normal subgroup N of G, show that G/N is abelian if and only if $[G,G] \subseteq N$.
5) Suppose $\varphi\colon G \to H$ is a homomorphism.
   (a) Show $\varphi([G,G]) = [\varphi(G), \varphi(G)]$.
   (b) Show that if H is abelian, then $[G,G] \subseteq \ker\varphi$.
6) Suppose H is a subgroup of G. Show $[H,H] \subseteq [G,G]$.

The commutator subgroup of G is sometimes also called the ***derived group*** of G, and denoted G' (like a derivative of G). Then the derived group of the derived group can be denoted G''. More generally, we can take the derived group of the derived group of the derived group of ... of G:

(3.1.10) **Definition.** For $i \in \mathbb{N}$, the ***ith derived group*** $\boldsymbol{G^{(i)}}$ is defined by induction:

i) $G^{(0)} = G$.

ii) $G^{(i+1)} = [G^{(i)}, G^{(i)}]$.

A normal subgroup of a normal subgroup need not be normal (see Exercise 3.1.16), but we have the following very important facts:

(3.1.11) **Exercise.**

1) A characteristic subgroup of a characteristic subgroup is characteristic. This means that if H char K char G, then H char G.
2) A characteristic subgroup of a normal subgroup is normal. That is, if H char $K \trianglelefteq G$, then $H \trianglelefteq G$.
3) $G^{(i)}$ char G for all i.

(3.1.12) **Definition.** G is ***solvable*** if $G^{(r)}$ is trivial for some $r \in \mathbb{N}$.

(3.1.13) **Exercise.** Show:

1) Every abelian group is solvable.
2) Subgroups of solvable groups are solvable.
3) Homomorphic images of solvable groups are solvable. (Equivalently, quotients of solvable groups are solvable.)
4) If N is a normal subgroup of G, such that N and G/N are solvable, then G is solvable.
5) There is a solvable group that is not abelian.

(3.1.14) **Definition.**

1) A ***subnormal series*** of G is a (finite) sequence $G_0, G_1, \ldots, G_r$ of subgroups of G, such that
$$G = G_0 \trianglerighteq G_1 \trianglerighteq G_2 \trianglerighteq \cdots \trianglerighteq G_r = \{1\}. \tag{3.1.15}$$
(That is, each G_i is normal in the preceding group G_{i-1}, and the series starts at G and ends at $\{1\}$.)
2) We say that (3.1.15) is a ***normal series*** of G if $G_i \trianglelefteq G$ for all i. (That is, G_i must be normal in all of G, not only in G_{i-1}.)
3) The ***quotients*** of the subnormal series (3.1.15) are the groups G_i/G_{i+1} for $0 \leq i < r$.

(3.1.16) **Exercise.** Show that a normal subgroup of a normal subgroup need not be a normal subgroup. (Therefore, the terms G_i in a subnormal series will typically <u>not</u> be normal subgroups of G.)

[*Hint:* One example can be obtained by letting $G = A_4$ be the subgroup consisting of all the even permutations in the symmetric group S_4, letting $H = \{(1,2)(3,4), (1,3)(2,4), (1,4)(2,3), e\}$ be the (unique) Sylow 2-subgroup of G, and letting $K = \langle (1,2)(3,4) \rangle$. Then $H \trianglelefteq G$ (in fact, H char G, because H consists of all the solutions of $x^2 = e$ in G) and $K \trianglelefteq H$ (because H is abelian), but $K \not\trianglelefteq G$ (because $(1,2)(3,4)$ is conjugate to $(1,3)(2,4)$).]

(3.1.17) **Example.** If G is solvable, then, by definition, we have $G^{(r)} = \{1\}$ for some r. So we have the normal series
$$G = G^{(0)} \trianglerighteq G^{(1)} \trianglerighteq G^{(2)} \trianglerighteq \cdots \trianglerighteq G^{(r-1)} \trianglerighteq G^{(r)} = \{1\}.$$
This is called the ***derived series*** of G.

Note that all of the quotients of the derived series are abelian. (For short, we say that "the derived series has abelian quotients.") This is the key to the following alternate characterizations of solvability, any of which could have been taken as the definition:

(3.1.18) **Proposition.** *The following are equivalent:*

1) *G is solvable.*
2) *$G^{(r)} = \{1\}$ for some r.*
3) *G has a normal series with abelian quotients.*
4) *G has a subnormal series with abelian quotients.*

Proof. Note that (1 $\Leftrightarrow$ 2) is the definition of "solvable," (2 $\Rightarrow$ 3) is Example 3.1.17 (plus the sentence that follows it), and (3 $\Rightarrow$ 4) is obvious, because every normal series is a subnormal series.

(4 $\Rightarrow$ 2) Suppose $G = G_0 \trianglerighteq G_1 \trianglerighteq G_2 \trianglerighteq \cdots \trianglerighteq G_r = \{1\}$ is a subnormal series with abelian quotients. We will prove, by induction, that $G^{(i)} \subseteq G_i$ for $0 \leq i \leq r$. Then, letting $i = r$ yields $G^{(r)} \subseteq G_r = \{1\}$, so $G^{(r)}$ is trivial, as desired.

The base case is the observation that $G^{(0)} = G = G_0$. For the induction step, assume $G^{(i)} \subseteq G_i$. Since the subnormal series has abelian quotients, we know G_i/G_{i+1} is abelian, so Exercise 3.1.9(4) tells us

$$[G_i, G_i] \subseteq G_{i+1}. \tag{3.1.19}$$

Therefore

$$\begin{aligned} G^{(i+1)} &= [G^{(i)}, G^{(i)}] && \text{(definition of } G^{(i+1)}\text{)} \\ &\subseteq [G_i, G_i] && \text{(induction hypothesis)} \\ &\subseteq G_{i+1} && \text{(see 3.1.19).} \end{aligned}$$

□

(3.1.20) **Definitions.**

1) G is ***perfect*** if $G = [G, G]$.
2) N is a ***minimal normal subgroup*** of G if $\{1\} \neq N \trianglelefteq G$, and N does not properly contain any nontrivial, normal subgroup of G.

(3.1.21) **Exercises.** Assume G is finite. Show:

1) G is solvable if and only if no nontrivial subgroup of G is perfect.
2) G is solvable if and only if every nontrivial quotient of G has a nontrivial, abelian, normal subgroup.
3) If G is solvable, then every minimal normal subgroup of G is abelian.

§3.2. Nilpotent groups and central series (advanced)

If $\{G_i\}$ is any normal series of G, then, by the Correspondence Theorem (1.4.12), we can view G_i/G_{i+1} as a (normal) subgroup of G/G_{i+1}. For G to be solvable, we require each G_i/G_{i+1} to be an abelian subgroup of G/G_{i+1}. For G to be "nilpotent," we make the much stronger requirement that G_i/G_{i+1} is in the centre of G/G_{i+1}:

(3.2.1) **Definition.**

1) A ***central series*** of G is a normal series
$$G = G_0 \trianglerighteq G_1 \trianglerighteq G_2 \trianglerighteq \cdots \trianglerighteq G_r = \{1\}, \tag{3.2.2}$$
such that $G_i/G_{i+1} \subseteq Z(G/G_{i+1})$ for $0 \leq i < r$.
2) G is ***nilpotent*** if it has a central series.

Since every subgroup of the centre is abelian, it is clear that nilpotent groups are solvable. The converse is not true, but every abelian group is nilpotent.

(3.2.3) **Exercises.** Show:

1) Every abelian group is nilpotent.
[*Hint:* $G \trianglerighteq \{1\}$ is a central series if G is abelian.]
2) The normal series (3.2.2) is a central series if and only if $[G, G_i] \subseteq G_{i+1}$ for $0 \leq i < r$.

3) Subgroups of nilpotent groups are nilpotent.
[*Hint:* Let $H_i = H \cap G_i$.]

4) Homomorphic images (and quotients) of nilpotent groups are nilpotent.
[*Hint:* For $\varphi\colon G \xrightarrow{\text{onto}} H$, let $H_i = \varphi(G_i)$.]

For finite groups, the following result provides several highly varied characterizations of nilpotence. The terminology and notation used in each part will be explained as we reach the proof of that part of the theorem.

(3.2.4) **Theorem.** *Assume G is finite. Then the following are equivalent:*

1) *G is nilpotent.*
2) *G has a central series.*
3) *G has no self-normalizing, proper subgroups.*
4) *Every maximal subgroup of G is normal.*
5) *Every maximal subgroup of G contains the commutator subgroup of G. In other words, $[G,G] \subseteq \Phi(G)$.*
6) *Every Sylow subgroup of G is normal.*
7) *G is a direct product of groups of prime-power order.*

Note that (1 $\Leftrightarrow$ 2) of Theorem 3.2.4 is given by Definition 3.2.1(2). The remainder of this section proves the equivalence of (2)–(7).

(2 $\Rightarrow$ 3) A central series implies there are no self-normalizing, proper subgroups.

(3.2.5) **Definition.** A subgroup H of G is said to be ***self-normalizing*** if $N_G(H) = H$. (In other words, H is its own normalizer.)

We will employ the following two straightforward observations:

(3.2.6) **Exercises.**

1) A subnormal series $G = G_0 \trianglerighteq G_1 \trianglerighteq G_2 \trianglerighteq \cdots \trianglerighteq G_r = \{1\}$ is a central series if and only if $[G, G_i] \subseteq G_{i+1}$ for $0 \leq i < r$.
2) Suppose H and K are subgroups of G. Show that $K \subseteq N_G(H)$ if and only if $[H,K] \subseteq H$.

(3.2.7) **Proposition.** *If G has a central series, then no proper subgroup of G is self-normalizing.*

Proof. Let H be a proper subgroup of G, and let $G = G_0 \trianglerighteq G_1 \trianglerighteq G_2 \trianglerighteq \cdots \trianglerighteq G_r = \{1\}$ be a central series of G. Since $H \neq G$, there is some i, such that $G_i \not\subseteq H$. By choosing i to be maximal, we may assume $G_{i+1} \subseteq H$. (Note that, since $1 \in H$, we must have $i < r$.) Then (using Exercise 3.2.6(1)) we have

$$[H, G_i] \subseteq [G, G_i] \subseteq G_{i+1} \subseteq H,$$

so Exercise 3.2.6(2) tells us $G_i \subseteq N_G(H)$. Since $G_i \not\subseteq H$, this implies $N_G(H) \neq H$. □

(3 $\Rightarrow$ 4) No self-normalizing subgroups implies maximal subgroups are normal.

(3.2.8) **Definition.** A subgroup M of G is ***maximal*** if M is proper, and is not contained in any larger, proper subgroup of G. That is, if H is a subgroup of G, such that $M \subseteq H \subseteq G$, then either $H = M$ or $H = G$.

(3.2.9) **Lemma.** *If G has no self-normalizing, proper subgroups, then every maximal subgroup of G is normal.*

Proof. Let M be a maximal subgroup of G. Obviously, we have $M \subseteq N_G(M) \subseteq G$. Since M is maximal, one of the two inclusions must be an equality. However, the leftmost inclusion cannot be an equality, since G has no self-normalizing, proper subgroups. Therefore, we must have $N_G(M) = G$, so $M \trianglelefteq G$. □

(4 ⇔ 5) Maximal subgroups are normal iff they contain the commutator subgroup.

(3.2.10) **Definition.** The ***Frattini subgroup*** of G is the intersection of all the maximal subgroups of G. It is denoted $\Phi(G)$.

Thus, by definition, a subgroup is contained in $\Phi(G)$ if and only if it is contained in every maximal subgroup of G.

(3.2.11) **Lemma.** *If $[G,G] \subseteq \Phi(G)$, then every maximal subgroup of G is normal.*

Proof. Let M be a maximal subgroup of G. By assumption, we have $[G,G] \subseteq \Phi(G)$ and, by definition, we have $\Phi(G) \subseteq M$. Therefore

$$[M,G] \subseteq [G,G] \subseteq \Phi(G) \subseteq M,$$

so $M \trianglelefteq G$ (see Exercise 3.2.6(2)). □

(3.2.12) **Proposition.** *If G is finite, and every maximal subgroup of G is normal, then $[G,G] \subseteq \Phi(G)$.*

Proof. Let M be a maximal subgroup of G. By assumption, we know that $M \trianglelefteq G$. Then combining the Correspondence Theorem (1.4.12) with the maximality of M implies that G/M has no nontrivial, proper subgroups. Therefore, G/M must be a cyclic group (of prime order). In particular, G/M is abelian, so $[G,G] \subseteq M$ (see Exercise 3.1.9(4)).

Since M is an arbitrary maximal subgroup of G, this implies that $[G,G]$ is contained in every maximal subgroup, so $[G,G]$ is contained in the intersection of all the maximal subgroups of G. In other words, $[G,G] \subseteq \Phi(G)$. □

(4 ⇒ 6) If maximal subgroups are normal, then Sylow subgroups are normal.

(3.2.13) **Lemma.** *Assume G is finite. Then every proper subgroup of G is contained in a maximal subgroup of G.*

Proof. Suppose some subgroup H is not contained in a maximal subgroup. (This will lead to a contradiction.) Then:

- H cannot be maximal (because H is contained in itself), so there is some subgroup H_1, such that $H \subsetneq H_1 \subsetneq G$.
- H_1 cannot be maximal (because H is contained in H_1), so there is some subgroup H_2, such that $H_1 \subsetneq H_2 \subsetneq G$.
- Continuing in this way, we inductively construct an infinite sequence $H_1, H_2, \ldots$ of subgroups of G, such that

$$H \subsetneq H_1 \subsetneq H_2 \subsetneq H_3 \subsetneq H_4 \subseteq \cdots.$$

This contradicts the fact that the finite group G has only finitely many subgroups, so it is impossible to construct an infinite sequence of distinct subgroups. □

(3.2.14) **Lemma.** *Assume*

- *P is a Sylow subgroup of a finite group G, and*
- *H is a subgroup of G that contains $N_G(P)$.*

Then H is self-normalizing.

Proof. Note: The method used in this proof is called the ***Frattini argument***.

Note that $P \subseteq H$ (because $P \subseteq N_G(P) \subseteq H$). Then, since $P \in \mathrm{Syl}_p(G)$, we have $P \in \mathrm{Syl}_p(H)$. (This is because any power of p that divides $|H|$ must also divide $|G|$.) Now, let $g \in N_G(H)$. Then ${}^gP \subseteq {}^gH = H$. Since conjugation by g is a bijection (indeed, it is an automorphism of G), we know

that ${}^gP$ has the same order as P, so this implies that ${}^gP$ is also a Sylow p-subgroup of H. So P and ${}^gP$ are two Sylow p-subgroups of H. Now, the key to the proof is to

apply Sylow's Conjugacy Theorem with the subgroup H in the role of G.

This tells us that the Sylow subgroups P and ${}^gP$ are conjugate **in H**. That is, there is some $h \in H$, such that ${}^h({}^gP) = P$. Then ${}^{hg}P = {}^h({}^gP) = P$, so $hg \in N_G(P) \subseteq H$. Multiplying on the left by h^{-1}, we see that $g \in h^{-1}H = H$. Since g is an arbitrary element of $N_G(H)$, this implies $N_G(H) \subseteq H$, so H is self-normalizing. □

(3.2.15) **Proposition.** *If some Sylow subgroup of G is not normal, then some maximal subgroup of G is not normal.*

Proof. Suppose $P \in \mathrm{Syl}_p(G)$, and $N_G(P) \neq G$. Then $N_G(P)$ is a proper subgroup, so it is contained in a maximal subgroup M (see Lemma 3.2.13). Therefore, M is a subgroup of G that contains $N_G(P)$, so Lemma 3.2.14 tells us $N_G(M) = M \neq G$, so $M \not\trianglelefteq G$. □

(3.2.16) **Exercises.** Here are two more applications of the Frattini Argument.

1) Assume G is finite. Show $\Phi(G)$ is nilpotent.
   [*Hint:* Let $P \in \mathrm{Syl}_p(\Phi(G))$. If P is not normal in G, then $N_G(P)$ is contained in a maximal subgroup of G. Use the Frattini argument to show that G is contained in this maximal subgroup.]
2) Assume G is finite. Show that if $G/\Phi(G)$ is nilpotent, then G is nilpotent.
   [*Hint:* Let $P \in \mathrm{Syl}_p(G)$, let $\overline{\ }: G \to G/\Phi(G)$ be the natural homomorphism, and suppose $N_G(P)$ is contained in a maximal subgroup M. Since $\overline{P} \trianglelefteq \overline{G}$ (why?), we know $P\Phi(G) \trianglelefteq G$. So the Frattini Argument implies that $M = G$, which is a contradiction.]

(6 ⇒ 7) Sylow subgroups normal implies a product of p-groups.

(3.2.17) **Exercise.** Suppose N_1 and N_2 are normal subgroups of G, such that $N_1N_2 = G$, and $N_1 \cap N_2 = \{1\}$. Show that the map $\varphi(n_1, n_2) = n_1n_2$ is an isomorphism from $N_1 \times N_2$ to G.

[*Hint:* Since $N_i \trianglelefteq G$, we have $[G, N_i] \subseteq N_i$. Therefore $[N_1, N_2] \subseteq N_1 \cap N_2 = \{1\}$. So every element of N_1 commutes with every element of N_2. Use this to show that φ is a homomorphism.]

(3.2.18) **Proposition.** *If every Sylow subgroup of G is normal, then G is a direct product of groups of prime-power order.*

Proof. Let us assume, for simplicity, that $|G| = p^aq^b$ has only two distinct prime factors. Let $P \in \mathrm{Syl}_p(G)$ and $Q \in \mathrm{Syl}_q(G)$. By assumption, we have $P \trianglelefteq G$ and $Q \trianglelefteq G$. Also, since $\gcd(|P|, |Q|) = 1$, Lagrange's Theorem implies that $P \cap Q = \{1\}$. Then, from Exercise 2.2.10(2), we have

$$|PQ| = \frac{|P|\,|Q|}{|P \cap Q|} = \frac{p^a\,q^b}{1} = |G|,$$

so $PQ = G$. Therefore $G \cong P \times Q$ (see Exercise 3.2.17). □

The proof of the general case (without assuming that $|G|$ has only two prime factors) relies on the following generalization of Exercise 3.2.17.

(3.2.19) **Exercise.** Suppose $N_1, \ldots, N_r$ are normal subgroups of G, such that

- $N_1N_2 \cdots N_r = G$, and
- for each i, we have $N_i \cap (N_1N_2 \cdots N_{i-1}\,N_{i+1}N_{i+2} \cdots N_r) = \{1\}$. (Note that N_i is the only factor missing from the product.)

Show the map $\varphi(n_1, n_2, \ldots, n_r) = n_1n_2 \cdots n_r$ is an isomorphism from $N_1 \times N_2 \times \cdots \times N_r$ to G.

(7 ⇒ 1) p-groups and their direct products are nilpotent.

(3.2.20) **Exercise.** If G and H are nilpotent, then $G \times H$ is nilpotent.
[*Hint:* Let $G = G_0 \trianglerighteq G_1 \trianglerighteq G_2 \trianglerighteq \cdots \trianglerighteq G_r = \{1\}$ and $H = H_0 \trianglerighteq H_1 \trianglerighteq H_2 \trianglerighteq \cdots \trianglerighteq H_s = \{1\}$ be central series. If $r = s$, then $G \times H = (G_0 \times H_0) \trianglerighteq (G_1 \times H_1) \trianglerighteq (G_2 \times H_2) \trianglerighteq \cdots \trianglerighteq (G_r \times H_r) = \{(1,1)\}$ is also a central series.]

(3.2.21) **Lemma.** *If $G/Z(G)$ is nilpotent, then G is nilpotent.*

Proof. Let $\overline{G} = G/Z(G)$, and let $\overline{G} = \overline{G_0} \trianglerighteq \overline{G_1} \trianglerighteq \overline{G_2} \trianglerighteq \cdots \trianglerighteq \overline{G_r} = \{1\}$ be a central series of $\overline{G}$. By the Correspondence Theorem, there is a normal subgroup G_i of G, such that $\overline{G_i} = G_i/Z(G)$, for each i.

To complete the proof, we show that $G = G_0 \trianglerighteq G_1 \trianglerighteq G_2 \trianglerighteq \cdots \trianglerighteq G_r \trianglerighteq \{1\} = \{1\}$ is a central series, by verifying the condition in Exercise 3.2.3(2). Let $\overline{\ }: G \to \overline{G}$ be the natural homomorphism. For $i < r$, we have

$$\begin{aligned} \overline{[G, G_i]} &= [\overline{G}, \overline{G_i}] && \text{(Exercise 3.1.9(5))} \\ &\subseteq \overline{G_{i+1}} && \text{(apply Exercise 3.2.3(2) to the central series } \{\overline{G_i}\}\text{),} \end{aligned}$$

so $[G, G_i] \subseteq G_{i+1}$. For $i = r$, we have $\{1\} = \overline{G_r} = G_r/Z(G)$, so $G_r = Z(G)$; therefore $[G, G_r] = \{1\}$. □

(3.2.22) **Warning.** Suppose $N \trianglelefteq G$. Unlike for solvable groups, knowing that N and G/N are nilpotent does not imply that G is nilpotent. (For example, it may be the case that N and G/N are abelian, but $[G, N] = N$.)

(3.2.23) **Proposition.** *Every p-group is nilpotent.*

Proof. Let P be a p-group. Since $\{1\}$ is obviously nilpotent (in fact, it is abelian), we may assume P is nontrivial. So $Z(P)$ is nontrivial (see Exercise 2.3.13(1)). Therefore $|P/Z(P)| < |P|$, so, by induction on the order of P, we may assume $P/Z(P)$ is nilpotent. Then Lemma 3.2.21 tells us that P is nilpotent. □

(3.2.24) **Corollary.** *Any direct product of (finitely many) groups of prime-power order is nilpotent.*

Proof. Suppose P_i is a p_i-group, for $i = 1, 2, \ldots, r$. From Proposition 3.2.23, we know that each P_i is nilpotent, so Exercise 3.2.20 (and induction on r) tells us that the direct product $P_1 \times P_2 \times \cdots \times P_r$ is nilpotent. □

§3.3. Lower central series and upper central series (optional)

Analogous to the derived series of a solvable group, we have two different natural central series of a nilpotent group.

(3.3.1) **Definition.**

1) We define $\gamma_i(G)$ inductively:
   (i) $\gamma_0(G) = G$.
   (ii) $\gamma_{i+1}(G)/ = [G, \gamma_i(G)]$.
   This is called the ***lower central series*** of G. (Alternatively, it is sometimes called the ***descending central series***.)
2) We define $Z_i(G)$ inductively:
   (i) $Z_0(G) = \{1\}$.
   (ii) $Z_{i+1}(G)/Z_i(G) = Z\big(G/Z_i(G)\big)$.
   This is called the ***upper central series*** of G. (Alternatively, it is sometimes called the ***ascending central series***.)

(3.3.2) **Exercise.** Show:

1) $\gamma_i(G)$ and $Z_i(G)$ are characteristic in G, for all $i \in \mathbb{N}$.

2) If $\gamma_s(G) = \{1\}$ for some s, then the normal series
$$G = \gamma_0(G) \trianglerighteq \gamma_1(G) \trianglerighteq \gamma_2(G) \trianglerighteq \cdots \trianglerighteq \gamma_{s-1}(G) \trianglerighteq \gamma_s(G) = \{1\}$$
is a central series.
3) If $Z_r(G) = G$ for some r, then the normal series
$$G = Z_r(G) \trianglerighteq Z_{r-1}(G) \trianglerighteq Z_{r-2}(G) \trianglerighteq \cdots \trianglerighteq Z_1(G) \trianglerighteq Z_0(G) = \{1\}$$
is a central series. (Note that the numbering here is in the opposite order from our usual numbering of the subgroups in a central series.)
4) If $G = G_0 \trianglerighteq G_1 \trianglerighteq G_2 \trianglerighteq \cdots \trianglerighteq G_r = \{1\}$ is any central series of G, then, for $0 \le i \le r$, we have
$$\gamma_i(G) \subseteq G_i \subseteq Z_{r-i}(G).$$
(Thus, $\gamma_i(G)$ provides a lower bound on G_i, and $Z_{r-i}(G)$ provides an upper bound. This is the reason they are called the lower and upper central series.)

This provides two additional characterizations of nilpotence:

(3.3.3) **Exercise.** Use Exercise 3.3.2(4) to prove the following are equivalent:
1) G is nilpotent.
2) $\gamma_r(G) = \{1\}$ for some $r \in \mathbb{N}$.
3) $Z_r(G) = G$ for some $r \in \mathbb{N}$.

(3.3.4) *Remark.* It follows from Exercise 3.3.2(4) that, for all $r \in \mathbb{N}$, we have $\gamma_r(G) = \{1\}$ if and only if $Z_r(G) = G$. If G is nilpotent, then
1) such an r does exist, and
2) the smallest such r is called the ***nilpotence class*** of G.

§3.4. Supersolvable groups (optional)

(3.4.1) **Definition.** G is ***supersolvable*** if it has a normal series with cyclic quotients.

(3.4.2) **Exercise.**
1) Show that every supersolvable group is solvable.
2) Show that every finite, nilpotent group is supersolvable.

(3.4.3) **Proposition.** *If G is supersolvable, then $[G, G]$ is nilpotent.*

The proof uses the following observation:

(3.4.4) **Lemma.** *If G is cyclic, then $\mathrm{Aut}(G)$ is abelian.*

Proof. We may assume $G = \mathbb{Z}_n$. For any $\varphi, \sigma \in \mathrm{Aut}(\mathbb{Z}_n)$, if we let $k = \varphi(1)$ and $\ell = \sigma(1)$, then we have $\varphi(x) = kx$ and $\sigma(x) = \ell x$ for all x. So
$$\varphi(\sigma(x)) = k(\ell x) = \ell(kx) = \sigma(\varphi(x)).$$ □

(3.4.5) *Remark.* The proof shows that $\mathrm{Aut}(\mathbb{Z}_n)$ is isomorphic to $\mathbb{Z}_n^\times$, the multiplicative group of invertible elements in the ring $\mathbb{Z}_n$.

Proof of Proposition 3.4.3. Let $G = G_0 \trianglerighteq G_1 \trianglerighteq G_2 \trianglerighteq \cdots \trianglerighteq G_r = \{1\}$ be a normal series with cyclic quotients. Defining $G_i' = G' \cap G_i$ yields a normal series for G', and we claim it is a central series. Since G normalizes both G_i and G_{i+1}, it acts by conjugation on the quotient G_i/G_{i+1}. This yields a homomorphism from G to $\mathrm{Aut}(G_i/G_{i+1})$. Since Lemma 3.4.4 tells us that $\mathrm{Aut}(G_i/G_{i+1})$ is abelian, we know that G' must be in the kernel of this homomorphism (see Exercise 3.1.9(5)), so G' centralizes G_i/G_{i+1}, which means $[G', G_i] \subseteq G_{i+1}$. Therefore
$$[G', G_i'] = [G', G' \cap G_i] \subseteq [G', G'] \cap [G', G_i] \subseteq G' \cap G_{i+1} = G_{i+1}'.$$ □

Here is another application of the same lemma:

(3.4.6) **Example.** If G'/G'' and G'' are cyclic, then $G'' = \{1\}$ (so G' is cyclic).

Proof. G acts on G'' by conjugation, yielding a homomorphism $G \to \mathrm{Aut}(G'')$. Since $\mathrm{Aut}(G'')$ is abelian (see Lemma 3.4.4), we know G' is in the kernel of this homomorphism (see Exercise 3.1.9(5)). So G' centralizes G''. This means that $G'' \subseteq Z(G')$. Since G'/G'' is cyclic, this implies $G'/Z(G')$ is cyclic. So G' is abelian (see Exercise 2.3.12(1)). Hence, its commutator subgroup G'' is trivial (see Exercise 3.1.9(2)). □

§3.5. Simple groups and composition series

(3.5.1) **Assumption.** Assume G is finite.

(3.5.2) **Definition.**

1) G is ***simple*** if the only normal subgroups of G are $\{1\}$ and G (and G is nontrivial).
2) A subnormal series is ***without repetition*** if all of the terms in the series are distinct: it is of the form $G = G_0 \rhd G_1 \rhd G_2 \rhd \cdots \rhd G_r = \{1\}$, where the notation $G_i \rhd G_{i+1}$ indicates that G_{i+1} is a proper, normal subgroup of G_i.

It is easy to see that the abelian simple groups are precisely the groups of prime order. The nonabelian simple groups are much more interesting.

(3.5.3) **Exercise.** Assume G is a nonabelian simple group. Show:

1) $[G, G] = G$.
2) G is not solvable.
3) The only subnormal series of G without repetition is $G = G_0 \rhd G_1 = \{1\}$.

(3.5.4) *Remark.* Recall that the ***alternating group*** is the group of all even permutations in the symmetric group S_n. Although we will not prove this, the alternating group A_5 (of order 60) is not solvable. In fact, it is simple, and every group of smaller order is solvable.

Any nontrivial group G has a subnormal series $G \rhd \{1\}$ without repetition. If G is not simple, then it has a nontrivial, proper, normal subgroup N, and we can insert this to make a longer subnormal series $G \rhd N \rhd \{1\}$. We call the new series a "refinement" of the original series.

(3.5.5) **Definitions.**

1) We say that a subnormal series $G = G_0^* \rhd G_1^* \rhd G_2^* \rhd \cdots \rhd G_s^* = \{1\}$ is a ***refinement*** of a subnormal series $G = G_0 \rhd G_1 \rhd G_2 \rhd \cdots \rhd G_r = \{1\}$ if $\{G_j^*\}_{j=0}^s$ can be obtained by inserting additional terms between some of the groups in the series $\{G_i\}_{i=0}^r$. More precisely, there is a sequence $0 = j_0 < j_1 < \cdots < j_r = s$, such that $G_{j_i}^* = G_i$, for $0 \le i \le r$.
2) Let $\{G_i\}_{i=0}^r$ be a subnormal series without repetition (with $G_0 = G$ and $G_r = \{1\}$).
   (a) This subnormal series is a ***composition series*** if every quotient G_i/G_{i+1} is a simple group.
   (b) Its ***length*** is the number r.

(3.5.6) **Exercise.** Every (finite) group has a composition series. In fact, show that every subnormal series without repetition has a refinement that is a composition series.

[*Hint:* Let $\{G_j^*\}_{j=0}^s$ be a refinement of $\{G_i\}_{i=0}^r$ of maximal length. (Why does such a refinement exist?) If $\{G_j^*\}_{j=0}^s$ is not a composition series, then there is some m, such that G_m^*/G_{m+1}^* is not simple, so it has a nontrivial, proper, normal subgroup. By the Correspondence Theorem, this means there is a normal subgroup N of G_m^* with $G_m^* \rhd N \rhd G_{m+1}^*$, so we can create a refinement of length $s+1$ by inserting N between G_m^* and G_{m+1}^*. This is a contradiction.]

A group may have many different composition series.

(3.5.7) **Example.** Here are two different composition series of the group $\mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_5$:

$$\begin{array}{ccccccc} \mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_5 & \rhd & \mathbb{Z}_2 \times \mathbb{Z}_3 \times \{0\} & \rhd & \mathbb{Z}_2 \times \{0\} \times \{0\} & \rhd & \{(0,0,0)\} \\ \mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_5 & \rhd & \{0\} \times \mathbb{Z}_3 \times \mathbb{Z}_5 & \rhd & \{0\} \times \{0\} \times \mathbb{Z}_5 & \rhd & \{(0,0,0)\} \end{array}$$

The quotients of the first composition series are (in order) $\mathbb{Z}_5, \mathbb{Z}_3, \mathbb{Z}_2$, and the quotients of the second series are the same groups, but in a different order: $\mathbb{Z}_2, \mathbb{Z}_3, \mathbb{Z}_5$.

(3.5.8) **Theorem** (Jordan-Hölder Theorem)**.** *All composition series of G have the same length, and have the same quotients* (*including multiplicity*)*, but perhaps in a different order.*

More precisely, suppose

$$G = G_0 \rhd G_1 \rhd G_2 \rhd \cdots \rhd G_r = \{1\} \quad \textit{and} \quad G = G_0^* \rhd G_1^* \rhd G_2^* \rhd \cdots \rhd G_s^* = \{1\}$$

are any two composition series of G. For $i \in \{1, 2, \ldots, r\}$ and $j \in \{1, 2, \ldots, s\}$, let $H_i = G_{i-1}/G_i$ and $H_j^* = G_{j-1}^*/G_j^*$. Then

1) *$r = s$, and*
2) *there is a permutation π of $\{1, 2, \ldots, r\}$, such that $H_i \cong H_{\pi(i)}^*$ for all i.*

The proof of the Jordan-Hölder Theorem is rather technical, so we will consider only some special cases:

(3.5.9) **Exercises.**

1) Suppose $G = G_0 \rhd H \rhd \{1\}$ and $G = G_0 \rhd K \rhd \{1\}$ are two composition series for G (of length two). Show that either:
   - $H = K$ (so it is obvious that the two composition series have the same quotients), or
   - $G/H \cong K$ and $H \cong G/K$.

   [*Hint:* Note that H, K, G/H, and G/K are simple. If $H \neq K$, then the simplicity implies $HK = G$ and $H \cap K = \{1\}$, so the map $(h, k) \mapsto hk$ is an isomorphism from $H \times K$ to G (see Exercise 3.2.17).]
2) More generally, prove the Jordan-Hölder Theorem in the special case where $s = 2$.

   [*Hint:* If $G_1 \neq G_1^*$, then combining the normality of G_1 with the simplicity of G_1^* and G/G_1^* implies $G \cong G_1 \times G_1^*$.]
3) Suppose $\{G_i\}_{i=0}^r$ and $\{G_j^*\}_{j=0}^s$ are two composition series of G. Show there is some i, such that $G_i/G_{i-1} \cong G/G_1^*$.

   [*Hint:* Choose i to be minimal with $G_{i+1}G_1^* \neq G$, and define a homomorphism $\Phi\colon G_i/G_{i+1} \to G/G_1^*$ by $\Phi(gG_{i+1}) = gG_1^*$. Show Φ is a well-defined isomorphism by exploiting the normality of G_{i+1} in G_i, the simplicity of G_i/G_{i+1} and G/G_1^*, and the fact that $G_iG_1^* = G$.]

(3.5.10) **Definition.** The quotients of a composition series of G are called the ***composition factors*** of G. (The Jordan-Hölder Theorem implies that these quotients do not depend on the choice of the composition series.)

(3.5.11) **Example.** The only composition series of the symmetric group S_5 is $S_5 \rhd A_5 \rhd \{1\}$, so the composition factors of S_5 are $\mathbb{Z}_2$ and A_5.

(3.5.12) **Exercise.** Show that G is solvable if and only if every composition factor of G is cyclic of prime order.

Chapter 4

Constructions of groups

The techniques in this chapter produce new examples of groups, and also provide useful descriptions of known groups.

§4.1. Informal look at groups defined by generators and relations

The following example illustrates that providing (sufficiently many) equations involving the generators of a group suffices to completely determine the group operation.

(4.1.1) **Example.** Suppose we know that $G = \langle x, y \rangle$ is generated by elements x and y, and we also know that $xy = y^{-1}x$. Then it is not difficult to see that $\{ x^i y^j \mid i, j \in \mathbb{Z} \}$ is closed under multiplication. More precisely,

$$(x^i y^j)(x^k y^\ell) = x^{i+k} y^{\epsilon j + \ell}, \text{ where } \epsilon \in \{\pm 1\}, \text{ with } \epsilon = 1 \text{ if and only if } k \text{ is even.} \tag{4.1.2}$$

This implies that every element of G can be written in the form $x^i y^j$, so (4.1.2) is a formula for the product of any two elements of G.

(4.1.3) **Definition** (informal)**.**

1) A ***relation*** between elements $x_1, \ldots, x_n$ of a group is an equation of the form
$$x_{i_1}^{p_1} x_{i_2}^{p_2} \cdots x_{i_r}^{p_r} = x_{j_1}^{q_1} x_{j_2}^{q_2} \cdots x_{j_s}^{q_s}.$$
(For example, $a^3 b^{-2} a = b^{-1} a^7 b$ is a relation between a and b.)
2) To define a group by ***generators and relations*** (or by giving a "***presentation***"), one provides
(a) a list $x_1, \ldots, x_n$ of generators of the group, and
(b) a list $R_1, \ldots, R_\ell$ of relations that are satisfied by the generators.
Then $\langle x_1, x_2, \ldots, x_n \mid R_1, R_2, \ldots, R_\ell \rangle$ is the group generated by $x_1, x_2, \ldots, x_n$, such that the given relations are all satisfied, and all other relations satisfied by the generators are consequences of $R_1, R_2, \ldots, R_\ell$.

(4.1.4) **Examples.**

1) For $n \in \mathbb{Z}^+$, $\langle x \mid x^n = 1 \rangle$ is the cyclic group of order n.
2) $\langle a, b \mid a^3 = b^5 = 1,\ ab = ba \rangle$ is (isomorphic to) $\mathbb{Z}_3 \times \mathbb{Z}_5$: the elements a and b are generators of $\mathbb{Z}_3$ and $\mathbb{Z}_5$, respectively, and the final relation tells us that these two subgroups commute with each other.
3) The group $\langle c \mid c^2 = c^3 = 1 \rangle$ is the trivial group. This is because the given relations imply $c = c^3 \cdot (c^2)^{-1} = 1 \cdot 1^{-1} = 1$. Thus, although the relation $c = 1$ was not given explicitly, it is a consequence of the given relations.

4) Every finite group G can be defined by generators and relations. For example, let $x_1, \ldots, x_n$ be a list of all the elements of G. For $1 \le i, j \le n$, there is some $m_{i,j}$, such that $x_i x_j = x_{m_{i,j}}$. Then
$$G = \langle\, x_1, \ldots, x_n \mid \{x_i x_j = x_{m_{i,j}}\}_{i,j=1}^{n} \,\rangle,$$
because the relations explicitly determine the entire multiplication table of the group G.

(4.1.5) *Remark.* Although, to simplify the notation, Definition 4.1.3 considers only finite generating sets and finite sets of relations, the idea generalizes to sets of any cardinality: the group $\langle X \mid \mathcal{R} \rangle$ can be defined for any set X of generators and any set $\mathcal{R}$ of relations between elements of X. In this setting, every group (not only the finite ones) can be defined by generators and relations.

(4.1.6) **Exercises.**

1) Show $|\langle\, x, y \mid x^3 = y^5 = 1,\ xy = y^2 x \,\rangle| \le 3$.
[*Hint:* Show $y = 1$ by using the fact that $x^3 = 1$ must commute with y.]
2) Show $\langle\, a, b \mid aba^{-1} = b^2,\ bab^{-1} = a^2 \,\rangle$ is trivial.
[*Hint:* $[a, b] = [b, a]^{-1}$.]
3) Show $|\langle\, r, s \mid r^7 = 1,\ rsr^{-1} = s^2 \,\rangle| \le 7 \cdot 127$.
[*Hint:* Show $\langle s \rangle$ is normal, so $\langle r \rangle \langle s \rangle$ is the entire group.]
4) Show $|\langle x, y, z \mid x^a = y^b = z^c = 1,\ xy = yxz,\ xz = zx,\ yz = zy \,\rangle| \le abc$, for all $a, b, c \in \mathbb{Z}^+$.

(4.1.7) **Example.** Show that $\langle\, a, b \mid a^{14} = b^{15} = 1,\ ab = ba,\ a^2 = b^3 \,\rangle \cong \mathbb{Z}_2 \times \mathbb{Z}_3$.

Proof. The group is abelian (because $ab = ba$), and additive notation will be more convenient, so let us write:
$$G = \langle\, a, b \mid 14a = 15b = 0,\ a + b = b + a,\ 2a = 3b \,\rangle.$$
Ignoring the relation $2a = 3b$ yields a group $H = \langle\, a, b \mid 14a = 15b = 0,\ a + b = b + a \,\rangle$. It is fairly clear that $H \cong \mathbb{Z}_{14} \times \mathbb{Z}_{15}$, with a and b corresponding to the generators $(1, 0)$ and $(0, 1)$ of $\mathbb{Z}_{14}$ and $\mathbb{Z}_{15}$, respectively. Now, G adds the additional relation that $2a = 3b$. So we want a group that is like H, except that $2a - 3b = 0$; in other words, $(2, -3) = 0$. Since $k \cdot 0 = 0$ for all $k \in \mathbb{Z}$, we must have $k(2, -3) = 0$; this means that every element of $\langle (2, -3) \rangle$ is 0. Now, quotient groups are exactly designed to do as little as is necessary to make the elements of a (normal) subgroup trivial, so we see that $G \cong H/\langle (2, -3) \rangle$. Since $\mathbb{Z}_{14}/\langle 2 \rangle \cong \mathbb{Z}_2$ and $\mathbb{Z}_{15}/\langle -3 \rangle \cong \mathbb{Z}_3$, it is not difficult to see that $(\mathbb{Z}_{14} \times \mathbb{Z}_{15})/\langle (2, -3) \rangle \cong \mathbb{Z}_2 \times \mathbb{Z}_3$. □

For abelian groups, as in the above example, adding an additional relation corresponds to modding out a cyclic subgroup. For nonabelian groups, most cyclic subgroups are not normal, and therefore cannot be modded out, but the following important result shows that additional relations do correspond to modding out a normal subgroup, or, equivalently, to taking a homomorphic image.

(4.1.8) **Theorem** (Von Dyck's Theorem)**.** *Assume*

- $G = \langle\, x_1, \ldots, x_n \mid R_1, R_2, \ldots, R_\ell \,\rangle$ *is a group defined by generators and relations,*
- $\hat{x}_1, \ldots, \hat{x}_n$ *are elements of a group H, and*
- *the relations $\hat{R}_1, \ldots, \hat{R}_\ell$ are satisfied in H, where $\hat{R}_j$ is obtained by replacing the symbols $x_1, \ldots, x_n$ in R_j with $\hat{x}_1, \ldots, \hat{x}_n$.*

Then there is a (unique) homomorphism $\varphi\colon G \to H$, such that $\varphi(x_i) = \hat{x}_i$ for all i.

A proof can be based on the following observation:

(4.1.9) **Exercise.** Suppose $a \in G$, and let N be the set of all elements of G that are products of conjugates of powers of a. That is:
$$N = \{\, {}^{g_1}(a^{k_1})\, {}^{g_2}(a^{k_2}) \cdots {}^{g_s}(a^{k_s}) \mid s \in \mathbb{N},\ g_1, \ldots, g_s \in G,\ k_1, \ldots, k_s \in \mathbb{Z} \,\}.$$
Show that N is the (unique) smallest normal subgroup of G that contains a.

[*Hint:* By definition, N is closed under products and inverses, and is invariant under conjugation, so it is a normal subgroup of G. Conversely, any normal subgroup of G that contains a must contain a and its inverse, and all their conjugates, and is closed under multiplication.]

Idea of proof. Write $H = \langle\, \hat{x}_1, \dots, \hat{x}_n \mid \widehat{\mathcal{R}} \,\rangle$, where $\widehat{\mathcal{R}}$ is some set of relations (see Example 4.1.4(4) and Remark 4.1.5). Since the relations $\hat{R}_1, \dots, \hat{R}_\ell$ are satisfied in H, we may assume that $\widehat{\mathcal{R}}$ contains $\hat{R}_1, \dots, \hat{R}_\ell$. To simplify the notation, let us assume that $\widehat{\mathcal{R}}$ contains only one additional relation:

$$H = \langle\, \hat{x}_1, \dots, \hat{x}_n \mid \hat{R}_1, \hat{R}_2, \dots, \hat{R}_\ell,\ \hat{R} \,\rangle.$$

By moving a term to the left side (that is, replacing the relation $A = B$ with $AB^{-1} = 1$), we may assume the right-hand side of the relation $\hat{R}$ is 1, so R is of the form $\hat{x}_{j_1}^{q_1} \hat{x}_{j_2}^{q_2} \cdots \hat{x}_{j_t}^{q_t} = 1$.

Let

- $a = x_{j_1}^{q_1} x_{j_2}^{q_2} \cdots x_{j_t}^{q_t} \in G$,
- N be the smallest <u>normal</u> subgroup of G that contains a,
- $\overline{x}_i = x_i N \in G/N$, and
- $\overline{R}_1, \dots, \overline{R}_n, \overline{R}$ be the relations that are obtained from $\hat{R}_1, \hat{R}_2, \dots, \hat{R}_\ell, \hat{R}$ by replacing $\hat{x}_1, \dots, \hat{x}_n$ with $\overline{x}_1, \dots, \overline{x}_n$.

Then the quotient $\overline{G} = G/N$ satisfies the relations $\overline{R}_1, \dots, \overline{R}_n$ (since G satisfies the relations $R_1, R_2, \dots, R_\ell$), and it also satisfies the relation $\overline{R}$, since $\overline{a} = 1$.

Conversely, we claim that every relation between $\overline{x}_1, \dots, \overline{x}_n$ is a consequence of $\overline{R}_1, \dots, \overline{R}_n, \overline{R}$, so

$$G/N \cong \langle\, \overline{x}_1, \dots, \overline{x}_n \mid \overline{R}_1, \overline{R}_2, \dots, \overline{R}_\ell,\ \overline{R} \,\rangle \overset{*}{\cong} \langle\, \hat{x}_1, \dots, \hat{x}_n \mid \hat{R}_1, \hat{R}_2, \dots, \hat{R}_\ell,\ \hat{R} \,\rangle = H.$$

(The isomorphism marked $*$ is obtained by simply changing the names of the variables: replacing each symbol $\overline{x}_i$ with $\hat{x}_i$.)

To prove the claim, suppose we have a relation $\overline{x}_{i_1}^{p_1} \overline{x}_{i_2}^{p_2} \cdots \overline{x}_{i_m}^{p_m} = 1$. This means that $x_{i_1}^{p_1} x_{i_2}^{p_2} \cdots x_{i_m}^{p_m} \in N$. So Exercise 4.1.9 tells us that

$$x_{i_1}^{p_1} x_{i_2}^{p_2} \cdots x_{i_m}^{p_m} = g_1(a^{k_1})\, g_2(a^{k_2}) \cdots g_s(a^{k_s})$$

is a product of conjugates of powers of a. By writing a and each g_i as a product of the generators of G, this equation expresses a relation between $x_1, \dots, x_n$, so, by the definition of G, this must be a consequence of the relations $R_1, R_2, \dots, R_\ell$. Therefore,

$$\overline{x}_{i_1}^{p_1} \overline{x}_{i_2}^{p_2} \cdots \overline{x}_{i_m}^{p_m} = \overline{g_1(a^{k_1})\, g_2(a^{k_2}) \cdots g_s(a^{k_s})} \tag{4.1.10}$$

is a consequence of the relations $\overline{R}_1, \overline{R}_2, \dots, \overline{R}_\ell$.

Furthermore, the relation $\overline{R} = 1$ says $\overline{a} = 1$, which implies that the right-hand side of (4.1.10) is trivial. Therefore, combining $\overline{R}$ with the relations $\overline{R_1}, \overline{R_2}, \dots, \overline{R_\ell}$ implies that $\overline{x}_{i_1}^{p_1} \overline{x}_{i_2}^{p_2} \cdots \overline{x}_{i_m}^{p_m} = 1$. Since this represents an arbitrary relation in G/N, we conclude that every relation in G/N is a consequence of $\overline{R_1}, \dots, \overline{R_n}, \overline{R}$, as claimed. □

Here is a useful consequence for finite groups:

(4.1.11) **Corollary.** *Suppose*

- $G = \langle\, x_1, \dots, x_n \mid R_1, R_2, \dots, R_\ell \,\rangle$ *is a group defined by generators and relations,*
- $H = \langle\, \hat{x}_1, \dots, \hat{x}_n \rangle$ *is a <u>finite</u> group that is generated by n elements, such that the relations $\hat{R}_1, \hat{R}_2, \dots, \hat{R}_\ell$ are satisfied, and*
- $|G| \le |H|$.

Then $G \cong H$.

Proof. Von Dyck's Theorem implies there is a homomorphism $\varphi\colon G \to H$, such that $\varphi(x_i) = \hat{x}_i$ for all i. Then $\varphi(G)$ contains the generating set $\{\hat{x}_1, \dots, \hat{x}_n\}$ of H, so φ is onto. This implies that $|H| \le |G|$. Since we have assumed that $|G| \le |H|$, this implies that $|G| = |H|$. Any surjection between finite sets of the same cardinality is a bijection, so φ is an isomorphism. □

(4.1.12) **Example.** We can identify the group $G = \langle x, y \mid x^3 = y^5 = 1,\ xy = y^2x \rangle$ of Exercise 4.1.6(1). To do this, let $H = \langle \hat{x} \rangle$ be a cyclic group of order 3 (so $\hat{x}^3 = 1$), and let $\hat{y} = 1 \in H$. Then it is obvious that H satisfies the relations

$$\hat{x}^3 = \hat{y}^5 = 1,\ \hat{x}\hat{y} = \hat{y}^2\hat{x},$$

and Exercise 4.1.6(1) tells us that $|G| \leq 3 = |H|$. So Corollary 4.1.11 implies that $G \cong H$ is a cyclic group of order 3.

(4.1.13) **Exercises.**

1) Verify parts (1) and (2) of Example 4.1.4.
2) Show $\langle g, h \mid g^4 = h^4 = 1,\ gh = hg,\ g^2 = h^2 \rangle \cong \mathbb{Z}_2 \times \mathbb{Z}_4$.
3) Show $\langle f, t \mid f^2 = t^5 = ftft = 1 \rangle \cong D_{10}$. (The dihedral group of order 10.)
4) Show $\langle p, q \mid p^3 = q^{11} = 1,\ pqp^{-1} = q^{-1} \rangle \cong \mathbb{Z}_3$.
5) Show that $\langle a, b \mid aba = bab \rangle$ is not abelian.
   [*Hint:* The relation is satisfied by the permutations $(1\,2)$ and $(2\,3)$ in S_3.]
6) Let $G = \langle r, s \mid r^3s = sr^3 \rangle$ and $H = \langle x, y \mid xy^6 = y^6x \rangle$. Show that if G is infinite, then H is infinite.

(4.1.14) *Remark.* Although sometimes useful, generators and relations are often very hard to work with. For example, there is a group $G = \langle x_1, \ldots, x_n \mid R_1, \ldots, R_\ell \rangle$ given by generators and relations, such that there is no algorithm (that is, it is impossible to write a computer program) that can always determine (for all choices of $i_1, \ldots, i_m$) whether the product $x_{i_1}x_{i_2} \cdots x_{i_m}$ is the trivial element of the group. In the terminology of combinatorial group theory, this means that the "word problem" for G is undecidable. (For more information, look up "word problem for groups" on *Wikipedia*.

§4.2. Free groups and the proof of Von Dyck's Theorem (advanced)

Before stating the formal definition of a group given by generators and relations, we explain the special case where there are no relations between the generators.

(4.2.1) **Definition** (informal)**.** Let X be a set. The ***free group*** on X is the group that is generated by the elements of X, and without any relations between the generators that are not forced by the rules of group theory.

(4.2.2) **Notation.**

1) Think of the elements of X as "symbols" $x_1, x_2, \ldots$ (or "letters" $a, b, c, \ldots$).
2) We let $X^{-1} = \{ x^{-1} \mid x \in X \}$. (This is another set of symbols, disjoint from the first.)
3) $(X \cup X^{-1})^* = \{ \text{finite words } a_1a_2 \ldots a_n \mid a_i \in X \cup X^{-1},\ n \in \mathbb{N} \}$. Note that the special case $n = 0$ yields the "empty word," which we will call "e."
4) Concatenation ("putting them end-to-end") is a binary operation on $(X \cup X^{-1})^*$:
   $$(a_1a_2 \cdots a_m)(b_1b_2 \cdots b_n) = a_1a_2 \cdots a_mb_1b_2 \cdots b_n.$$
   Note that the identity element of this operation is e.
5) Two words w_1 and w_2 are equivalent ($w_1 \sim w_2$) if one can be obtained from the other by a finite sequence of insertions and/or deletions of words of the form xx^{-1} or $x^{-1}x$, where $x \in X$. It is not difficult to show that this is an equivalence relation.

(4.2.3) **Example.** Let $X = \{a, b, c\}$. The word $abb^{-1}a^2b$ is equivalent to the word $a^2c^{-1}cab$, because we can get from the first to the second by deleting bb^{-1}, and then inserting $c^{-1}c$.

(4.2.4) **Definition.** The ***free group*** on X is $(X \cup X^{-1})^*/\sim$, the set of equivalence classes of $(X \cup X^{-1})^*$ under the above equivalence relation.

(4.2.5) **Fact.** *The free group on X is a group:*

1) *If $w_1 \sim w_1'$ and $w_2 \sim w_2'$, then $w_1 w_2 \sim w_1' w_2'$ (for all $w_1, w_1', w_2, w_2' \in (X \cup X^{-1})^*$), so concatenation provides a well-defined binary operation on $(X \cup X^{-1})^*/\sim$.*
2) *It is clear that concatenation is associative.*
3) *The identity element is e, the empty word.*
4) *Every word has an inverse: $(a_1 a_2 \cdots a_n)^{-1} = a_n^{-1} a_{n-1}^{-1} \cdots a_1^{-1}$.*

(4.2.6) **Notation.** F_n is the free group on a set with n elements (usually $\{x_1, x_2, \ldots, x_n\}$). It may be called "the free group on n letters."

Note that if a word w contains a generator next to its inverse (that is, a subword of the form xx^{-1} or $x^{-1}x$), then deleting this subword yields an equivalent word that is shorter than the original one. Deleting more subwords yields shorter and shorter words, which cannot go on forever (since the original word had finite length), so we must eventually obtain a word with no such subword:

(4.2.7) **Fact.** *Every word is equivalent to a unique* ***reduced word****. That is, a word in which no generator is adjacent to its inverse.* (*For example, $abcb^{-1}$ is reduced, but $abb^{-1}c$ is not reduced.*)

(4.2.8) *Remark.* The uniqueness in Fact 4.2.7 implies that it is easy to determine whether two words represent the same element of the free group: simply put them in reduced form and check whether the two reduced words are equal. (On the other hand, Remark 4.1.14 implies that in some other groups it can be impossibly difficult to tell whether two products are equal.)

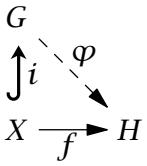
We have the following special case of Von Dyck's Theorem (4.1.8):

(4.2.9) **Theorem** (universal property of free groups)**.** *Suppose G is the free group on X, and H is any group. Then every function $f\colon X \to H$ extends to a unique homomorphism $\varphi\colon G \to H$.*

(4.2.10) *Remark.* In modern mathematical notation, the above theorem can be stated in terms of a "commutative diagram."

- We have the natural inclusion map $i\colon X \hookrightarrow G$.
- We are given a function $f\colon X \to H$.

That provides two sides (not the dashed one) of the following triangle:



The theorem says there is a (unique) homomorphism φ from G to H, the third side of the triangle (drawn dashed, because it is provided by the theorem, not given at the start), such that the diagram "commutes," which means that going from X to H via f (one side of the triangle) yields the same result as taking the other route, by going from X to H via G. That is, f is equal to the composition $\varphi \circ i$.

Idea of proof. In order for φ to extend f, we must have $\varphi(x) = f(x)$ for all $x \in X$. Then, since φ is a homomorphism, we must have $\varphi(x^{-1}) = f(x)^{-1}$. Hence, for any word $w = x_1^{\epsilon_1} x_2^{\epsilon_2} \cdots x_n^{\epsilon_n} \in (X \cup X^{-1})^*$, we must have

$$\varphi(x_1^{\epsilon_1} x_2^{\epsilon_2} \cdots x_n^{\epsilon_n}) = f(x_1)^{\epsilon_1} f(x_2)^{\epsilon_2} \cdots f(x_n)^{\epsilon_n}. \tag{4.2.11}$$

This proves that the homomorphism φ is unique (if it exists).

To prove existence, we take (4.2.11) as the definition of φ. It is clear that this map respects multiplication. The problem is to show that φ is well-defined: if $w_1 \sim w_2$, then $\varphi(w_1) = \varphi(w_2)$.

The formal proof requires some care, but the idea is easy enough:

$$\begin{aligned}\varphi(x_1x_2\cdots x_k a a^{-1} y_1 y_2 \cdots y_\ell) &= f(x_1)f(x_2)\cdots f(x_k)\ f(a)f(a)^{-1}\ f(y_1)f(y_2)\cdots f(y_\ell)\\ &= f(x_1)f(x_2\cdots f(x_k)f(y_1)f(y_2)\cdots f(y_\ell)\\ &= \varphi(x_1x_2\cdots x_k y_1 y_2\cdots y_\ell).\end{aligned}$$

□

(4.2.12) **Corollary.** *Every group is a homomorphic image of a free group.*

Proof. Let X be any set that generates G, and let F be the free group on X. Then the inclusion $X \hookrightarrow G$ extends to a homomorphism $\varphi\colon F \to G$. Since $\varphi(F)$ contains $\varphi(X) = f(X) = X$, which generates G, we must have $\varphi(F) = G$. So φ is a homomorphism from F onto G, so G is a homomorphic image of F. □

Here is another way of saying the same thing:

(4.2.13) **Corollary.** *Every group is* (*isomorphic to*) *a quotient of a free group. In other words, every group is isomorphic to F/N, for some free group F and some normal subgroup N of F.*

This observation is the foundation of the formal definition of a group defined by generators and relations.

(4.2.14) **Example.** To construct the group $G = \langle\, a, b \mid ab = b^{-1}a \,\rangle$, we first let F be the free group on $\{a, b\}$. Then F has the correct generators, but does not have the necessary relation between a and b. To impose this relation, we pass to a quotient group in which the relation does hold. Namely, if we rewrite the desired relation as $aba^{-1}b = 1$ (by moving all terms to the left-hand side), then we see that we are looking for a quotient F/N of F in which $aba^{-1}b$ is trivial. This means that we need to have $aba^{-1}b \in N$.

However, we want to make sure that the group F/N does not satisfy any relations that are not consequences of the required relation $aba^{-1}b = 1$. That means that we do not want to make anything equal to the identity if it does not have to be equal to the identity: we want as few elements of F to be trivial as possible. So N should be as small as possible. Hence, we let N be the smallest normal subgroup of F that contains $aba^{-1}b$. (Since the intersection of normal subgroups is a normal subgroup, this smallest normal subgroup is unique.)

The same idea works in general, and yields the following official construction of a group defined by ***generators and relations***.

(4.2.15) **Definition.** Suppose X is a set, and $\{w_1, v_1, \ldots, v_n, v_n\}$ is a set of words in $(X \cup X^{-1})^*$. Then

$$\langle\, X \mid w_1 = v_1,\ \ldots, w_n = v_n \,\rangle = F/N,$$

where

- F is the free group on X, and
- N is the (unique) smallest normal subgroup of F that contains $w_i v_i^{-1}$ for all i.

(4.2.16) **Exercise.** Use Definition 4.2.15 (and the Correspondence Theorem) to provide a short proof of Von Dyck's Theorem (4.1.8).

§4.3. Semidirect products (optional)

(4.3.1) **Recall.**

1) If A and B are groups, then the Cartesian product $A \times B$ is a group under componentwise multiplication: $(a_1, b_1) \cdot (a_2, b_2) = (a_1a_2, b_1b_2)$. This group is called the ***direct product*** of A and B.

2) If A and B are normal subgroups of G, such that $AB = G$ and $A \cap B = \{1\}$, then $G \cong A \times B$, and we say that G is the ***internal direct product*** of A and B.

If A and B are abelian, then $A \times B$ is abelian. This section describes an important generalization of the direct product that often yields nonabelian groups, even when the factors A and B are abelian. The key is to remove the assumption that B is normal, while retaining the assumption that A is normal:

(4.3.2) **Definition.** If A and B are subgroups of G, such that

$$A \trianglelefteq G, \ AB = G, \text{ and } A \cap B = \{1\},$$

then we say that G is the ***internal semidirect product*** of A and B.

If A, B, and G are as in Definition 4.3.2, then, since A is normalized by B, there is an action of B on A by conjugation: ${}^b a = bab^{-1}$. Note that, for all $a_1, a_2 \in A$ and $b_1, b_2 \in B$, we have

$$(a_1 b_1)(a_2 b_2) \ = a_1 (b_1 a_2 b_1^{-1}) b_1 b_2 = \ a_1 \, {}^{b_1}a_2 \, b_1 \, b_2. \tag{4.3.3}$$

Furthermore, if we let $\varphi_b(a) = {}^b a$, then $\varphi_b \in \mathrm{Aut}(A)$. Indeed, φ is a homomorphism from B to $\mathrm{Aut}(A)$. Conversely, any homomorphism from B to $\mathrm{Aut}(A)$ arises from a semidirect product:

(4.3.4) **Definition.** Assume A and B are groups.

1) Recall from Exercise 2.1.3 that an action of B on A is defined by any homomorphism $\varphi\colon B \to S_A$. We say that the action is ***by automorphisms*** if $\varphi_b \in \mathrm{Aut}(A)$, for all $b \in B$. (In other words, if $\varphi\colon B \to \mathrm{Aut}(A)$.) We usually write ${}^b a$ for $\varphi_b(a)$.
2) For any action of B on A by automorphisms, we define a binary operation on $A \times B$ by:

$$(a_1, b_1) * (a_2, b_2) = (a_1 \, {}^{b_1}a_2, \ b_1 b_2).$$

The resulting group is called the ***semidirect product*** of A and B, and is denoted $A \rtimes B$.

(4.3.5) *Remarks.*

1) The binary operation in Definition 4.3.4(2) is based on the formula in (4.3.3).
2) If φ is the trivial homomorphism, then $A \rtimes B$ is the direct product $A \times B$.
3) If the choice of the homomorphism φ is not clear from the context, it can be included in the notation as a subscript: $A \rtimes_\varphi B$.
4) The notation $A \rtimes B$ is used because it combines the symbol for a direct product ("$\times$") with the symbol for a normal subgroup — the right half of $\rtimes$ is $\triangleleft$. This is intended to reflect the fact that A is normal ("$\triangleleft$") in $A \rtimes B$.

(4.3.6) **Exercises.** Let $G = A \rtimes B$, with respect to some action of B on A by automorphisms.

1) Verify that the binary operation $*$ in Definition 4.3.4 satisfies the group axioms.
2) Show:
   (a) $\widehat{A} = \{ (a, 1) \mid a \in A \}$ is a subgroup of G that is isomorphic to A.
   (b) $\widehat{B} = \{ (1, b) \mid b \in B \}$ is a subgroup of G that is isomorphic to B.
   (c) G is the internal semidirect product of $\widehat{A}$ and $\widehat{B}$.

(4.3.7) **Exercises.**

1) Show that the ***Klein bottle group*** $\langle x, y \mid yxy^{-1} = x^{-1} \rangle$ is isomorphic to a semidirect product $\mathbb{Z} \rtimes \mathbb{Z}$.
2) Show that the symmetric group S_4 is the semidirect product of a group of order 4 and a group of order 6.

(4.3.8) **Exercises** (contrast with Exercise 2.3.12(3))**.** Assume p and q are prime, with $q \equiv 1 \pmod{p}$.

1) Show there is a nonabelian group of order pq.

   [*Hint:* Cauchy's Theorem implies that $\mathrm{Aut}(\mathbb{Z}_q)$ has a subgroup P of order p. Then $\mathbb{Z}_q \rtimes P$ is a nonabelian group of order pq.]

2) Show that every group of order pq is the internal semidirect product of a subgroup of order q and a subgroup of order p.
3) Show the group in (1) is unique (up to isomorphism).

[*Hint:* You may assume (without proof) that $\mathrm{Aut}(\mathbb{Z}_q)$ is cyclic, and therefore has a unique subgroup of order p.]

(4.3.9) **Exercise.** Recall that a natural number is ***square-free*** if no divisor of the number (other than 1) is a perfect square.

It is known that every group of square-free order is solvable. Assuming this, show that every group of square-free order is the semidirect product of two cyclic groups.

[*Hint:* Use Example 3.4.6 to show that G'' is trivial. Every abelian group of square-free order is cyclic.]

Part II

Rings and Modules

Chapter 5

Summary of undergraduate ring theory

(5.0.1) **Notation.** R is always a ring (with 1).

§5.1. Elementary facts and definitions

(5.1.1) **Definition.** A ***ring*** is a set R together with two operations, called addition ($+$) and multiplication ($\cdot$), such that:

1) $(R, +)$ is an (additive) abelian group, with identity element 0.
2) There exists a multiplicative identity element. This means there exists $1 \in R$, such that $1 \cdot r = r \cdot 1 = r$ for all $r \in R$.
3) Multiplication is associative. This means $r(st) = (rs)t$ for all $r, s, t \in R$.
4) The distributive laws hold. More precisely, multiplication distributes over addition, both on the left and on the right. This means $r(s + t) = rs + rt$ and $(r + s)t = rt + st$ for all $r, s, t \in R$.

(5.1.2) **Warning.** Any two elements of a ring can be added, subtracted, or multiplied, but there is no requirement that they can be divided (even if they are nonzero). For example, $\mathbb{Z}$ is a ring, but the quotient of two elements of $\mathbb{Z}$ is usually not in $\mathbb{Z}$.

(5.1.3) *Other conventions.* Some textbooks do not require rings to have a multiplicative identity element, so they omit (2) from the definition.

(5.1.4) **Example.** Some rings to keep in mind are:

1) $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$,
2) $\mathbb{Z}_n = \{\overline{0}, \overline{1}, \overline{2}, \ldots, \overline{n-1}\}$,
3) polynomial rings, such as $\mathbb{R}[x]$, $\mathbb{C}[x]$, $\mathbb{Z}[x]$, and $\mathbb{Z}_n[x]$ (see Definition 5.4.2), and
4) the ring $\mathrm{Mat}_{n\times n}(\mathbb{C})$ of $n \times n$ matrices with entries in $\mathbb{C}$.

(5.1.5) **Exercises.** Let r, s, and t be elements of R. Prove the following facts from middle-school algebra:

1) $r \cdot 0 = 0 \cdot r = 0$,
2) $(-1) \cdot r = r \cdot (-1) = -r$,
3) $(-r) \cdot s = r \cdot (-s) = -rs$,
4) $(-r)(-s) = rs$,
5) $r(s - t) = rs - rt$ and $(r - s)t = rt - st$.

(5.1.6) **Warning.** In elementary school, you are taught that if $rs = 0$, then either $r = 0$ or $s = 0$. This is not always true for a general ring. (For example, $\overline{2} \cdot \overline{3} = \overline{0}$ in $\mathbb{Z}_6$.)

(5.1.7) **Definition** (ideals)**.**

1) A nonempty subset I of R is a ***left ideal*** if it is closed under addition and is also closed under multiplication on the left by elements of R. This means that
   - $i_1 + i_2 \in I$ for all $i_1, i_2 \in I$, and
   - $ri \in I$, for all $r \in R$ and $i \in I$.
2) Similarly, a nonempty subset I of R is a ***right ideal*** if it is closed under addition and closed under multiplication on the right by elements of R.
3) A subset of R that is both a left ideal and a right ideal is called an ***ideal*** (or, for emphasis, it can be called a ***two-sided ideal***). We write $I \trianglelefteq R$ to denote that I is a two-sided ideal of R.
4) The obvious (two-sided) ideals of R are $\{0\}$ and R. We call $\{0\}$ the ***zero*** ideal. (All other ideals are ***nonzero***.) A (right or left or two-sided) ideal I of R is said to be ***proper*** if $I \neq R$.
5) If I is an ideal of R, then the ***quotient ring*** R/I is the ring whose elements are the additive coset $r + I$ of I (with $r \in R$) with addition and multiplication defined by:
   - $(r_1 + I) + (r_2 + I) = (r_1 + r_2) + I$, and
   - $(r_1 + I)(r_2 + I) = r_1 r_2 + I$.

(5.1.8) **Exercises.**

1) Show that if I and J are (left or right or two-sided) ideals of R, then $I \cap J$ is a (left or right or two-sided) ideal of R.
2) Show that the intersection of any number of (left or right or two-sided) ideals of R is a (left or right or two-sided) ideal of R.
3) Show that if I and J are (left or right or two-sided) ideals of R, then $I + J$ is a (left or right or two-sided) ideal of R.
4) For any $x \in R$, show that the set $Rx = \{ rx \mid r \in R \}$ is the smallest left ideal that contains x. (Similarly, $xR = \{ xr \mid r \in R \}$ is the smallest right ideal that contains x.)

(5.1.9) **Basic facts** (commutative rings)**.**

1) A ring R is ***commutative*** if $rs = sr$ for all $r, s \in R$.
2) The rings $\mathbb{Z}$, $\mathbb{Z}_n$, $\mathbb{R}[x]$, $\mathbb{C}[x]$, $\mathbb{Z}[x]$, and $\mathbb{Z}_n[x]$ are commutative.
3) If $n > 1$, then $\mathrm{Mat}_{n\times n}(\mathbb{C})$ is not commutative. For example, we have
$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \neq \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$
4) In a commutative ring, there is no difference between left ideals and right ideals — all of them are two-sided.
5) Quotients of commutative rings are commutative: if R is a commutative ring, and I is an ideal of R, then R/I is commutative.

(5.1.10) **Basic facts.** Let R be a ring.

1) A subset S of R is a ***subring*** if it is a ring under the operations obtained by restricting the operations of G. Equivalently: for all $s_1, s_2, s \in S$, we have $s_1 + s_2 \in S$, $s_1 \cdot s_2 \in S$, $-s \in S$, and $1 \in S$.
2) A subring S of R is said to be *proper* if $S \neq R$.
3) Every subring of a commutative ring is commutative.
4) The intersection of any number of subrings of R is a subring of R.
5) The ***direct sum*** of two rings R and S is denoted $R \oplus S$. This is the ring whose elements are the elements of the Cartesian product $R \times S$, with addition and multiplication defined componentwise: $(r_1, s_2) + (r_2, s_2) = (r_1 + s_1, r_2 + s_2)$ and $(r_1, s_2) \cdot (r_2, s_2) = (r_1 s_1, r_2 s_2)$.

6) Direct sums of commutative rings are commutative: if R and S are commutative rings, then $R \oplus S$ is commutative.

§5.2. Homomorphisms and isomorphisms

(5.2.1) **Definition.** Assume R and S are rings.

1) A function $\varphi\colon R \to S$ is a (ring) ***homomorphism*** if it respects the ring operations. That is, we have
$$\varphi(r_1 + r_2) = \varphi(r_1) + \varphi(r_2), \qquad \varphi(r_1 r_2) = \varphi(r_1) \cdot \varphi(r_2), \qquad \text{and} \qquad \varphi(1_R) = 1_S$$
for all $r_1, r_2 \in R$. (Here, to avoid confusion, we have used 1_R to denote the multiplicative identity element of R, and 1_S to denote the multiplicative identity element of S.)
2) A bijective homomorphism is called an ***isomorphism***.
3) We say that R and S are ***isomorphic*** (and write $R \cong S$) if there is an isomorphism from R to S. This is an equivalence relation:
$$R \cong R, \qquad R \cong S \Rightarrow S \cong R, \qquad \text{and} \qquad R \cong S \cong T \Rightarrow R \cong T\,.$$
4) $\ker\varphi = \varphi^{-1}(0_S) = \{\, r \in R \mid \varphi(r) = 0_S \,\}$ is called the ***kernel*** of φ.

(5.2.2) **Exercises.** Assume φ is a homomorphism from R to S. Prove the following basic facts.

1) The inverse of an isomorphism is an isomorphism: if φ is an isomorphism, then φ^{-1} is an isomorphism from S to R.
2) Homomorphisms respect multiples and powers: we have
$$\varphi(kr) = k\varphi(r) \quad \text{and} \quad \varphi(r^\ell) = \varphi(r)^\ell$$
for $r \in R$, $k \in \mathbb{Z}$, and $\ell \in \mathbb{Z}^+$. (Also, $\varphi(0_R) = 0_S$.)
3) If T is a subring of R, then $\varphi(T)$ is a subring of S.
4) φ is one-to-one if and only if $\ker\varphi = \{0\}$.
5) $\ker\varphi$ is an ideal of R.
6) Conversely, if I is any ideal of R, then the function $\varphi(x) = x + I$ is a homomorphism from R to R/I whose kernel is I.

Isomorphisms preserve all properties that can be expressed in ring-theoretic terms. For example:

(5.2.3) **Examples.** Assume $\varphi\colon R \xrightarrow{\cong} S$, and let $r, r' \in R$. Then:

1) r commutes with r' if and only if $\varphi(r)$ commutes with $\varphi(r')$.
2) R is commutative if and only if S is commutative.
3) every left ideal of R is a two-sided ideal if and only if every left ideal of S is a two-sided ideal.
4) R has a nonzero, proper ideal if and only if S has a nonzero, proper ideal.
5) R is a PID (or UFD, or Euclidean domain) if and only if S is a PID (or UFD, or Euclidean domain) (see Definition 5.3.2).
6) etc.

Although fundamental, the material in the remainder of this section is **optional**, because none of it will be needed in later chapters.

(5.2.4) **Proposition** (1st, 2nd, and 3rd Isomorphism Theorems)**.**

1) *If φ is a homomorphism from R to S, then $R/\ker\varphi \cong \varphi(R)$.*
More precisely, an isomorphism $\overline{\varphi}\colon R/\ker\varphi \to \varphi(R)$ can be obtained by defining $\overline{\varphi}(r + \ker\varphi) = \varphi(r)$.
2) *Suppose $I \trianglelefteq R$ and S is a subring of R. Then*
(a) *$S + I$ is a subring of R (where $S + I = \{\, s + i \mid s \in S,\ i \in I \,\}$),*

(b) *$S \cap I$ is an ideal of S, and*
(c) *$(S+I)/I \cong S/(S \cap I)$.*

3) *Suppose I and J are ideals of R, with $I \subseteq J$. Then J/I is an ideal of R/I, and $(R/I)/(J/I) \cong R/J$.*

(5.2.5) **Proposition** (Correspondence Theorem)**.** *Suppose I is a two-sided ideal of R.*

1) *There is a one-to-one correspondence between the subrings of R that contain I and the subrings of R/I. Namely, the subring of R/I corresponding to a subring S of R is S/I.*
2) *There is a one-to-one correspondence between the* (*right, left, or two-sided*) *ideals of R that contain I and the* (*right, left, or two-sided*) *ideals of R/I. Namely, the ideal of R/I corresponding to an ideal J of R is J/I.*

§5.3. PIDs, UFDs, and Euclidean domains

After giving the necessary definitions, this section proves that

$$\text{Euclidean} \overset{5.3.5}{\Longrightarrow} \text{PID} \overset{5.3.11}{\Longrightarrow} \text{UFD}.$$

(5.3.1) **Assumption.** The ring R is **commutative**.

(5.3.2) **Definitions.**

1) An element r of R is a ***zero divisor*** if $r \neq 0$ and there exists $s \in R \smallsetminus \{0\}$, such that $rs = 0$.
2) R is an ***integral domain*** if it does not have any zero divisors.
3) R is a ***Euclidean domain*** if
   (a) R is an integral domain, and
   (b) there is a function $d\colon R \to \mathbb{N}$, such that
      - $d(r) = 0 \iff r = 0$, and
      - for all $a, b \in R$ with $b \neq 0$, there exists $q, r \in R$ with $a = bq + r$ and $d(r) < d(b)$.

   (The natural number $d(r)$ is often called the ***norm*** of the element r.)
4) An ideal I of R is ***principal*** if $I = Rr$, for some $r \in R$.
5) R is a ***principal ideal domain*** (or ***PID***, for short) if
   (a) R is an integral domain, and
   (b) every ideal of R is principal.
6) An element u of R is a ***unit*** if it has a multiplicative inverse (that is, there exists $r \in R$, such that $ur = 1$).
7) Two nonzero elements r and s are ***associates*** if $r = us$, for some unit u. (So the associates of 1 are precisely the units.) This is an equivalence relation.
8) We write $r \mid s$ to mean that r is a **divisor** of s (that is, there exists $t \in R$ with $rt = s$).
9) A nonzero element π of R is ***irreducible*** if it is not a unit, and every divisor of π is either a unit or an associate of π.
10) R is a ***unique factorization domain*** (or ***UFD***, for short) if every nonzero nonunit of R can be written uniquely as a product of irreducibles (up to associates and a permutation of the factors). More precisely, for every $r \in R$, such that $r \neq 0$ and r is not a unit:
    (a) there exist irreducibles $\pi_1, \pi_2, \ldots, \pi_k$, for some k, such that $r = \pi_1, \pi_2, \ldots, \pi_k$, and
    (b) if $r = \pi'_1, \pi'_2, \ldots, \pi'_\ell$, where $\pi'_1, \pi'_2, \ldots, \pi'_\ell$ are irreducible, then $k = \ell$, and there exists $\sigma \in S_k$, such that π_i is an associate of $\pi'_{\sigma(i)}$ for all i.
11) A nonzero element p of R is ***prime*** if it is not a unit and, for all $r, s \in R$ with $p \mid rs$, we have either $p \mid r$ or $p \mid s$. (Note that every prime element of an integral domain is irreducible, and that the converse is true in any UFD.)
12) The ***zero ring*** is the ring with only one element. Any ring with more than one element is a ***nonzero ring***.
13) If F is a nonzero commutative ring, and every nonzero element of F is a unit, then we say that F is a ***field***.

(5.3.3) **Examples.** For our purposes, the two most important examples of Euclidean domains are:

1) $\mathbb{Z}$ is a Euclidean domain: let $d(n) = |n|$.
2) If F is any field, then we will see in Section 5.4 that the polynomial ring $F[x]$ is a Euclidean domain, with $d(f(x)) = 1 + \deg f(x)$ (except that $d(0) = 0$).

We require the following basic fact:

(5.3.4) **Exercise.** Show that $\mathbb{N}$ is ***well-ordered***. (This means that every nonempty subset of $\mathbb{N}$ has a smallest element.)

[*Hint:* Prove by induction on n: If S is a subset of $\mathbb{N}$ that contains an element $\leq n$, then S has a smallest element.]

(5.3.5) **Proposition.** *Every Euclidean domain is a PID.*

Proof. Let I be any ideal of R. Since $\{0\} = R0$ is principal, we may assume $I \neq \{0\}$. Then $\{\, d(i) \mid i \in I \smallsetminus \{0\} \,\}$ is a nonempty subset of $\mathbb{Z}^+$, so it has a smallest element (because $\mathbb{N}$ is well-ordered; see Exercise 5.3.4). Hence, we may choose some $i_0 \in I$, such that

$$d(i_0) \leq d(i) \text{ for all nonzero } i \in I. \tag{5.3.6}$$

We claim that $I = Ri_0$. To see this, let i be an arbitrary element of I. Since R is Euclidean, there exist $q, r \in R$, such that $i = qi_0 + r$, and $d(r) < d(i_0)$. Now $r = i - qi_0 \in I - RI \subseteq I$, so we have $r \in I$ and $d(r) < d(i_0)$. Therefore, the minimality of $d(i_0)$ implies $r = 0$. So $i = qi_0 + r = qi_0 + 0 = qi_0 \in Ri_0$. Since i is an arbitrary element of I, this implies $I \subseteq Ri_0$. The opposite inclusion is obvious, because $i_0 \in I$ and I is an ideal. □

(5.3.7) **Example.** The fact that every Euclidean domain is a PID provides us with two important examples of PIDs:

- $\mathbb{Z}$ is a PID.
- If F is any field, then the polynomial ring $F[x]$ is a PID.

In Chapter 7, we will study the structure of "modules" over any PID, but these two rings are the most important examples.

Our proof that every PID is a UFD relies on the following important notion:

(5.3.8) **Definition.** R has the ***ascending chain condition*** (or ***acc***, for short) if it satisfies any of the following equivalent conditions:

1) Every ascending chain of ideals is eventually constant. That is,
   if $I_1 \subseteq I_2 \subseteq \cdots$ is any ascending chain of ideals,
   then there is some n, such that $I_n = I_{n+1} = I_{n+2} = \cdots$.
2) There is no infinitely long *strictly* increasing chain of ideals. That is,
   there do not exist ideals $I_1, I_2, \ldots$, such that $I_1 \subsetneq I_2 \subsetneq I_3 \subsetneq \cdots$.
3) Every ideal is finitely generated. That is, if I is any ideal, then there exist $x_1, x_2, \ldots, x_n \in R$ (for some n) such that $I = \langle x_1, \ldots, x_n \rangle$ is the ideal generated by $x_1, x_2, \ldots, x_n$.
4) Every nonempty set $\mathcal{I}$ of ideals has a maximal element. That is, there is some $M \in \mathcal{I}$, such that there does not exist $I \in \mathcal{I}$ with $M \subsetneq I$.

(*Alternate terminology:* saying that R is ***Noetherian*** is the same as saying that R has the acc.)

(5.3.9) **Exercises.** Assume R is an integral domain.

1) Show that if $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ is an increasing chain of ideals, then $\bigcup_{n=1}^{\infty} I_n$ is an ideal.
2) It is fairly obvious that the first two formulations in Definition 5.3.8 are equivalent.
   (a) Show that 5.3.8(3) is equivalent to 5.3.8(2).

   [*Hint:* ($\Leftarrow$) If $I_1 \subsetneq I_2 \subsetneq I_3 \subsetneq \cdots$, then $\bigcup_n I_n$ is an ideal that is not finitely generated (because any finite subset of I is contained in some I_n). ($\Rightarrow$) If I is not finitely generated, then a sequence $\{x_1, x_2, \ldots\}$ can be constructed by induction, such that $x_{n+1} \notin \langle x_1, \ldots, x_n \rangle =: I_n$.]

   (b) Show that 5.3.8(4) is equivalent to 5.3.8(2).

   [*Hint:* ($\Rightarrow$) If $\mathcal{I}$ has no maximal element, you can inductively construct a strictly increasing chain of ideals in $\mathcal{I}$. ($\Leftarrow$) If $\{I_n\}$ is a strictly increasing chain of ideals, then it has no maximal element.]

3) For $r, s \in R$, show that if $Rrs = Rr$, then either $r = 0$ or s is a unit.
4) Assume R has the acc. Show that if r is any element of R that is not a unit, then r is divisible by some irreducible.

   [*Hint:* Suppose r_0 is not divisible by any irreducible. By induction, construct a sequences $\{r_n\}_{n\geq 0}$ and $\{s_n\}_{n\geq 1}$, such that $r_n = r_{n+1}s_{n+1}$, and neither r_n nor s_n is a unit. Then $Rr_0 \subsetneq Rr_1 \subsetneq Rr_2 \cdots$.]
5) Show that if R has the acc, then every nonzero, non-unit element of R is the product of finitely many irreducibles.

   [*Hint:* Suppose r_0 is not the product of finitely many irreducibles. Use Exercise 5.3.9(4) to inductively construct sequences $\{\pi_n\}_{n\geq 1}$ and $\{r_n\}_{n\geq 0}$, such that $r_n = \pi_{n+1}r_{n+1}$. Then $Rr_0 \subsetneq Rr_1 \subsetneq Rr_2 \cdots$.]
6) A ring has the ***descending chain condition*** (or ***dcc***, for short) if every descending chain of ideals is eventually constant. Show that a nonzero integral domain with the dcc must be a field.

   [*Hint:* Given $r \in R$, you need to show that r has a multiplicative inverse. Consider the ideals $Rr, Rr^2, Rr^3, \ldots$.]

We now know that every element of a ring with the acc is a product of irreducibles. The following result is the key to proving that the product is unique when the ring is a PID.

(5.3.10) **Lemma.** *If R is a PID, then every irreducible element of R is prime.*

Proof. Assume $\pi, r_1, r_2 \in R$, such that π is irreducible, and neither r_1 nor r_2 is divisible by π. For each i, since R is a PID, we have $\langle r_i, \pi\rangle = Rd_i$ some $d_i \in R$. Then $\pi \in Rd_i$, so $d_i \mid \pi$. Since π is irreducible, this implies that we may assume (after multiplying by a unit) that d_i is either 1 or π. However, since $\pi \nmid r_i$, we have $r_i \notin R\pi$, so $d_i \neq \pi$. Therefore $d_i = 1$, which means $\langle r_i, \pi\rangle = R$. Hence, there exist $a_i, b_i \in R$, such that $a_i r_i + b_i\pi = 1$, so $a_i r_i \equiv 1 \pmod{\pi}$. Then

$$(r_1r_2)a_1a_2 = (a_1r_1)(a_2r_2) \equiv 1 \cdot 1 = 1 \not\equiv 0 \pmod{\pi},$$

so $r_1r_2 \not\equiv 0 \pmod{\pi}$. □

(5.3.11) **Proposition.** *Every PID is a UFD.*

Proof. We already saw in Exercise 5.3.9(5) that every nonzero non-unit in R is a product of irreducibles, so all that remains to prove is uniqueness.

Suppose $\pi_1 \cdots \pi_k = u\pi'_1 \cdots \pi'_\ell$, where every π_i and π'_j is irreducible, and u is a unit. The left-hand side is obviously divisible by π_k, so the right-hand side must also be divisible by π_k. From Lemma 5.3.10 (and induction), this implies that some π'_j is divisible by π_k. By permuting the factors on the right-hand side, we may assume it is π'_ℓ that is divisible by π_k. Since π'_ℓ and π_k are irreducible, this implies that they are associates, so $\pi'_\ell = u'\pi_k$, for some unit u'. Then we have

$$\pi_1 \cdots \pi_k = u\pi'_1 \cdots \pi'_\ell = u\pi'_1 \cdots \pi'_{\ell-1}(u'\pi_k) = uu'\pi'_1 \cdots \pi'_{\ell-1}\pi_k,$$

so $\pi_1 \cdots \pi_{k-1} = uu'\pi'_1 \cdots \pi'_{\ell-1}$. Then, by induction on k, we must have $k - 1 = \ell - 1$ (so $k = \ell$), and there exists $\sigma \in S_{k-1}$, such that π_i is an associate of $\pi'_{\sigma(i)}$ for all i. □

(5.3.12) **Exercises.**

1) Suppose $r, s \in R$. We say that d is a ***greatest common divisor*** (or ***gcd*** for short) of r and s if
   - d is a common divisor of r and s (that is, we have $d \mid a$ and $d \mid s$),
   - every common divisor of r and s is a divisor of d.

   Show that if R is a UFD, then any two elements of R have a greatest common divisor, and that the gcd is unique up to multiplying by a unit.
2) Assume R is a PID. Then R is a UFD, so any two elements have a gcd. Show:
   (a) $\langle r, s\rangle = \langle \gcd(r, s)\rangle$.
   (b) There exist $a, b \in R$, such that $ar + bs = \gcd(r, s)$.
3) Assume R is a PID, and $r \in R \smallsetminus \{0\}$. Show the ideal $\langle r\rangle$ is maximal if and only if r is irreducible. (An ideal I of R is ***maximal*** if it is a proper ideal, and there does not exist an ideal J, such that $I \subsetneq J \subsetneq R$.)

4) Show that the subring $\mathbb{Z} + \mathbb{Z}\sqrt{-6}$ of $\mathbb{C}$ is **not** a PID.
[*Hint:* You may assume (without proof) that $\mathbb{Z}+\mathbb{Z}\sqrt{-6}$ is a subring of $\mathbb{C}$. It is not a UFD, because 10 can be factored in two different ways.]

§5.4. Fields and polynomials

(5.4.1) **Exercise.**
1) Show that a nonzero commutative ring is a field if and only if it has no nonzero, proper ideals.
2) Show that a ring R is the zero ring if and only if $0_R = 1_R$.

(5.4.2) **Definition** (polynomials)**.** Assume F is a field.
1) A ***polynomial*** over F is any expression of the form
$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0,$$
where $n \in \mathbb{N}$ and each a_i is an element of F.
2) Adding leading zeroes does not change a polynomial: if $a_{n+1} = a_{n+2} = \cdots = a_{n+k}$, then the polynomial $\sum_{i=0}^{n+k} a_i x^i$ is considered to be equal to the polynomial $\sum_{i=0}^{n} a_i x^i$.
3) We call $\sum_{i=0}^{n} a_i x^i$ the ***zero polynomial*** if $a_i = 0$ for all i.
4) The ***degree*** of a polynomial $f(x) = \sum_{i=0}^{n} a_i x^i$ is the largest value of i, such that $a_i \neq 0$. (By convention, we may say that the degree of the zero polynomial is -1.) It is denoted by $\deg f(x)$. Note that $\deg f(x) \geq 0$ if $f(x)$ is not the zero polynomial.
5) A ***linear*** polynomial is a polynomial of degree 1.
6) The set of all polynomials over F is a commutative ring. It is called the ***ring of polynomials over F***, and is denoted $F[x]$. Addition and multiplication of polynomials are defined just as in high school. Namely, if
$$f(x) = \sum_{i=0}^{m} a_i x^i \quad \text{and} \quad g(x) = \sum_{i=0}^{n} b_i x^i,$$
then
$$f(x) + g(x) = \sum_{i=0}^{\max(m,n)} (a_i + b_i)x^i$$
and
$$f(x)\,g(x) = \sum_{i=0}^{m+n} c_i x^i, \quad \text{where} \quad c_k = \sum_{i=0}^{k} a_i b_{k-i}$$
(and where we use the convention that $a_i = 0$ for $i > m$ and $b_i = 0$ for $i > n$).
7) For $f(x) = \sum_{i=0}^{n} a_i x^i \in F[x]$ and $c \in F$, we let $f(c) = \sum_{i=0}^{n} a_i c^i \in F$. Note that the mapping $f(x) \mapsto f(c)$ is a ring homomorphism:
   - If $f(x) + g(x) = h(x)$, then $f(c) + g(c) = h(c)$.
   - If $f(x)\,g(x) = h(x)$, then $f(c)\,g(c) = h(c)$.
8) Let $c \in F$ and $f(x) \in F[x]$. We say that c is a ***root*** (or ***zero***) of $f(x)$ if $f(c) = 0$.

(5.4.3) **Exercise.** Assume F is a field, and let $f(x), g(x) \in F[x]$.
1) Show that $\deg\big(f(x) + g(x)\big) \leq \max\big(\deg f(x), \deg g(x)\big)$.
2) Show that if neither $f(x)$ nor $g(x)$ is the zero polynomial, then
$$\deg\big(f(x)\,g(x)\big) = \deg f(x) + \deg g(x).$$
Conclude that $f(x)\,g(x)$ is not the zero polynomial.

Polynomials can be defined over any ring, not only fields, but the following important result does not hold for most other rings:

(5.4.4) **Theorem.** *If F is a field, then $F[x]$ is a Euclidean domain.*

More precisely, we may let $d(f(x)) = 1 + \deg f(x)$. This is an immediate consequence of the following result:

(5.4.5) **Theorem** (Division Algorithm for Polynomials)**.** *Let $f(x), g(x) \in F[x]$, where F is a field. If $g(x)$ is not the zero polynomial, then there exist unique polynomials $q(x), r(x) \in F[x]$, such that*

$$f(x) = g(x)\,q(x) + r(x) \quad \textit{and} \quad \deg r(x) < \deg g(x).$$

Proof. We first prove the existence of $q(x)$ and $r(x)$. This is accomplished by induction on $\deg f(x)$. For the base case, note that if $\deg f(x) < \deg g(x)$, then we may let $q(x) = 0$ and $r(x) = f(x)$. Therefore, we now assume that $\deg f(x) > \deg g(x)$. (In particular, $f(x)$ is not the zero polynomial.) Write $f(x) = \sum_{i=0}^{m} a_i x^i$ and $g(x) = \sum_{i=0}^{n} b_i x^i$ with $a_m \neq 0$ and $b_n \neq 0$. Now, let

$$f'(x) = f(x) - \frac{a_m}{b_n} x^{m-n}\, g(x).$$

(It is important to note that $a_m/b_n \in F$, since $b_n \neq 0$ and F is a field. So $\frac{a_m}{b_n} x^{m-n}\, g(x)$ is the product of two polynomials, and is therefore in $F[x]$, because $m - n = \deg f - \deg g > 0$.) Note that, since $\deg f(x) = m$ and

$$\deg\left(\frac{a_m}{b_n} x^{m-n}\, g(x)\right) = \deg \frac{a_m}{b_n} x^{m-n} + \deg g(x) = (m-n) + n = m,$$

we have $\deg f'(x) \leq m$. However, we also have

$$\begin{aligned} \frac{a_m}{b_n} x^{m-n}\, g(x) &= \frac{a_m}{b_n} x^{m-n}(b_n x^n + (\textit{terms of degree} < n)) \\ &= \frac{a_m}{b_n} x^{m-n} \cdot b_n x^n + \frac{a_m}{b_n} x^{m-n} \cdot (\textit{terms of degree} < n)) \\ &= a_m x^m + (\textit{terms of degree} < m), \end{aligned}$$

so the leading term of $\frac{a_m}{b_n} x^{m-n}\, g(x)$ exactly cancels the leading term of $f(x)$. Therefore $\deg f'(x) < \deg f(x)$, so, by our induction hypothesis, there exist $q'(x), r(x) \in F[x]$, such that

$$f'(x) = g(x)\,q'(x) + r(x) \quad \text{and} \quad \deg r'(x) < \deg g(x).$$

Now, let $q(x) = \frac{a_m}{b_n} x^{m-n} + q'(x)$. Then

$$\begin{aligned} g(x)\,q(x) + r(x) &= g(x)\big(\tfrac{a_m}{b_n} x^{m-n} + q'(x)\big) + r(x) \\ &= \tfrac{a_m}{b_n} x^{m-n}\, g(x) + \big(g(x)\,q'(x) + r(x)\big) \\ &= \big(f(x) - f'(x)\big) + f'(x) \\ &= f(x). \end{aligned}$$

Now, to prove uniqueness, suppose that

$$f(x) = g(x)\,q(x) + r(x) \quad \text{and} \quad f(x) = g(x)\,q'(x) + r'(x),$$

with $\deg r(x), \deg r'(x) < \deg g(x)$. Then high-school algebra tells us

$$r(x) - r'(x) = g(x)\big(q'(x) - q(x)\big). \tag{5.4.6}$$

If $q'(x) \neq q(x)$, then $q'(x) - q(x)$ is not the zero polynomial, so we see from Exercise 5.4.3(2) that

$$\deg\big(r(x) - r'(x)\big) = \deg g(x) + \deg\big(q'(x) - q(x)\big) \geq \deg g(x).$$

This is a contradiction, since

$$\deg\big(r(x) - r'(x)\big) \leq \max\big(\deg r(x), \deg r'(x)\big) < \deg g(x).$$

So we must have $q'(x) = q(x)$. Then $q'(x) - q(x) = 0$, so (5.4.6) implies $r(x) = r'(x)$. We now know that $q'(x) = q(x)$ and $r'(x) = r(x)$, which establishes the desired uniqueness. □

Note that if the linear polynomial $x - c$ is a factor of $f(x)$ (that is, if there exists $q(x) \in F[x]$, such that $f(x) = (x - c)\,q(x)$), then

$$f(c) = (c - c)\,q(c) = 0 \cdot q(c) = 0,$$

so c is a root of $f(x)$. The converse is a consequence of the Division Algorithm:

(5.4.7) **Corollary.** *Suppose $c \in F$ and $f(x) \in F[x]$, where F is a field. If c is a root of $f(x)$, then there exists $q(x) \in F[x]$, such that $f(x) = (x - c)\, q(x)$.*

Proof. From the Division Algorithm (5.4.5), we know there exist $q(x), r(x) \in F[x]$, such that $f(x) = (x - c)q(x) + r(x)$, and $\deg r(x) < \deg(x - c)$. Since $\deg(x - c) = 1$, this implies $\deg r(x) \in \{0, -1\}$, so $r(x)$ is some constant a. (If $\deg r(x) = -1$, then $a = 0$.) This implies $r(t) = a$ for all $t \in F$, so, in particular, we have $r(c) = a$. Then

$$0 = f(c) = (c - c)q(c) + r(c) = 0 \cdot q(c) + a = a.$$

This implies $r(x) = a = 0$, so we conclude that $r(x)$ is the zero polynomial. Therefore

$$f(x) = (x - c)q(x) + r(x) = (x - c)q(x) + 0 = (x - c)q(x). \qquad \square$$

(5.4.8) **Definition.** Let $c \in F$ and $f(x) \in F[x]$, where F is a field. From Corollary 5.4.7, we know that c is a root of $f(x)$ if and only if the linear polynomial $x - c$ is a factor of $f(x)$. We say that c is a ***repeated root*** of $f(x)$ if $(x - c)^2$ is a factor of $f(x)$.

Chapter 6

Modules over a ring

A vector space has objects, called vectors, that can be added and can be multiplied by scalars. In a first course in linear algebra, the scalars are usually real numbers (or perhaps complex numbers). Certainly, the scalars come from a field. We will now study the situation in which scalars are allowed to come from any ring. However, the term “vector space" is reserved for cases where the scalars are from a field; the space is called a “module” when the vectors come from a more general ring.

§6.1. Definition and basic facts

(6.1.1) **Definition.** A (left) ***module*** over R (or, for short, an ***R-module***) is an abelian group M (written additively), together with an operation of scalar multiplication $R \times M \to M$ that satisfies the following natural axioms for all $r, s \in R$ and $m, n \in M$:

1) (distributivity) $r(m+n) = rm + rn$ and $(r+s)m = rm + sm$.
2) (associativity) $r(sm) = (rs)m$.
3) (identity axiom) $1m = m$ (where 1 is the multiplicative identity element of R).

(6.1.2) *Other conventions.* In some situations, it is more natural to put the scalars on the right, instead of on the left (writing mr instead of rm). Then M is called a right R-module, and the axioms are:

$$(m+n)r = mr + nr, \quad m(r+s) = mr + ms, \quad (mr)s = m(rs), \quad m1 = m.$$

(6.1.3) **Exercises.** Assume $r \in R$ and $m \in M$. Let 0_R and 0_M be the additive identity elements of R and M, respectively. Show:

1) $r\,0_M = 0_M$
2) $0_R\,m = 0_M$
3) $(-r)m = r(-m) = -(rm)$
4) If $rm = 0_M$, and r is a unit in R (that is, if r has a multiplicative inverse), then $m = 0_M$.

(6.1.4) **Examples.**

1) If $R = \mathbb{R}$ (or more generally, if R is any field), then the R-modules are precisely the vector spaces over R (that is, the vector spaces whose scalars are the elements of the field R).
2) Every (additive) abelian group A is a $\mathbb{Z}$-module. Namely, the axioms are satisfied by the usual multiplication of an integer times an element of the group (such as $3a = a + a + a$ and $-3a = -(a + a + a)$).
3) R is an R-module under left multiplication. (That is, the scalar multiplication is the same as the multiplication that is part of the structure of R as a ring.) This is often called the ***regular*** R-module.

4) More generally, every left ideal of R is an R-module.
5) For any $n \in \mathbb{Z}^+$, let R^n be the set of all n-tuples with entries in R. This is an R-module under the usual componentwise scalar multiplication of vectors: $r(s_1, \ldots, s_n) = (rs_1, \ldots rs_n)$.
6) Let $\mathrm{Mat}_{n\times n}(R)$ be the set of all $n \times n$ matrices with entries in R. This is a ring with the usual (entrywise) addition and (row-times-column) multiplication of matrices. If we identify R^n with the set $\mathrm{Mat}_{n\times 1}(R)$ of n-dimensional column vectors with entries in R, then matrix multiplication provides R^n with the structure of a $\mathrm{Mat}_{n\times n}(R)$-module.

(6.1.5) **Exercises.**

1) Assume M is an R-module, and $\varphi\colon S \to R$ is a ring homomorphism (with $\varphi(1_S) = 1_R$). Show that M is an S-module, where the scalar multiplication is defined by $s \cdot m = \varphi(s)\, m$.
2) Suppose
   - V is a vector space over a field F, and
   - $T\colon V \to V$ is a linear transformation.

   Show that V is an $F[x]$-module, where the scalar multiplication is defined by
$$f(x) \cdot v = \sum_{i=0}^{n} a_i\, T^i v \qquad \text{if } f(x) = \sum_{i=0}^{n} a_i x^i \text{ with } a_i \in F.$$
3) Find an example of an R-module M, such that we have $rm = 0$ for some $r \in R \smallsetminus \{0\}$ and $m \in M \smallsetminus \{0\}$. (This complements Exercise 6.1.3(4) and demonstrates that not all properties of vector spaces remain true for modules.)
   [*Hint:* Take $R = \mathbb{Z}$.]

§6.2. Submodules, quotients, homomorphisms, and annihilators

(6.2.1) **Notation.** In this section, M is always an R-module.

(6.2.2) **Definitions.**

1) A subset N of M is a ***submodule*** of M if N is an R-module under the same operations as M. Equivalently, N is not empty, and is closed under addition and scalar multiplication: for all $n_1, n_2, n \in N$ and $r \in R$, we have
$$n_1 + n_2 \in N \quad \text{and} \quad rn \in N.$$
   We write $N \leq M$ to denote that N is a submodule of M.
2) The obvious submodules of M are $\{0\}$ (the ***trivial submodule***) and M. A submodule that is not all of M is said to be a ***proper*** submodule of M.
3) Suppose N is a submodule of M. Then N is a subgroup of the additive group of M, so
   - we have the left coset $m + N$ for each $m \in M$, and
   - M/N is the set of all left cosets of N.

   It is straightforward to verify that M/N is a module, under the natural operations
$$(m_1 + N) + (m_2 + N) = (m_1 + m_2) + M \quad \text{and} \quad r(m + N) = rm + N,$$
   for $m_1, m_2, m \in M$ and $r \in R$. This is called the ***quotient*** of M by N.

(6.2.3) *Remark.* Unlike in group theory, where only the normal subgroups give rise to a quotient group, every submodule of M gives rise to a quotient module.

(6.2.4) **Examples.**

1) Considering R as an R-module under left-multiplication, the submodules of R are precisely the left ideas of R.
2) If A is a $\mathbb{Z}$-module, then the submodules of A are precisely the subgroups of A.
3) If R is a field, and V is a vector space over R, then the submodules of V are precisely the vector subspaces of V.

(6.2.5) **Exercises.**

1) Assume $N_1, N_2, N_3 \leq M$. Show:
   (a) $N_1 + N_2 \leq M$, where $N_1 + N_2 = \{ n_1 + n_2 \mid n_1 \in N_1,\ n_2 \in N_2 \}$.
   (b) If $N_1 \subseteq N_3$, then $(N_1 + N_2) \cap N_3 = N_1 + (N_2 \cap N_3)$.
2) Assume $s \in R$, and R is commutative.
   (a) Show $sM \leq M$, where $sM = \{ sm \mid m \in M \}$.
   (b) Show $\{ m \in M \mid sm = 0 \} \leq M$.
3) Suppose I is a right ideal of R. Define
$$I^{\perp} = \{ m \in M \mid im = 0 \text{ for all } i \in I \}.$$
   Show $I^{\perp}$ is a submodule of M.

Linear transformations are the maps between vector spaces that respect the operations of addition and scalar multiplication. For modules (as for groups, rings, and most other algebraic objects), the maps that respect the operations are called homomorphisms:

(6.2.6) **Definitions.** Assume M and N are R-modules, and let $\varphi\colon M \to N$.

1) φ is a ***homomorphism*** if $\varphi(m_1 + m_2) = \varphi(m_1) + \varphi(m_2)$ and $\varphi(rm) = r\,\varphi(m)$ for all $m_1, m_2, m \in M$ and $r \in R$.
2) A bijective homomorphism is called an ***isomorphism***.
3) We say that M is ***isomorphic*** to N (and write $M \cong N$) if there exists an isomorphism from M to N. It is straightforward to check that this is an equivalence relation.
4) If φ is a homomorphism, then:
   (a) $\ker \varphi = \{ m \in M \mid \varphi(m) = 0 \}$. This is the ***kernel*** of φ.
   (b) $\varphi(M) = \{ \varphi(m) \mid m \in M \}$. This is the ***image*** of φ.

(6.2.7) **Exercises.** Assume $\varphi\colon M \to N$ is a homomorphism of R-modules.

1) Show that both the kernel and the image of φ are submodules. More precisely, show $\ker \varphi \leq M$ and $\varphi(M) \leq N$.
2) Show that the image of a submodule is a submodule: if $M_1 \leq M$, then $\varphi(M_1) \leq N$.
3) Show that the inverse image of a submodule is a submodule: if $N_1 \leq N$, then $\varphi^{-1}(N_1) \leq M$.
4) Show that the sum of two homomorphisms is a homomorphism: if $\psi\colon M \to N$ is another homomorphism, then defining $(\varphi+\psi)(m) = \varphi(m)+\psi(m)$ yields another homomorphism from M to N.
5) φ is one-to-one if and only if $\ker \varphi = \{0\}$.

(6.2.8) **Exercise.** Assume M is an R-module, and $N \leq M$. Show that the natural map from M to M/N is a surjective homomorphism whose kernel is N.

(6.2.9) **Exercise.** Assume that M is an R-module, and that R is a commutative. Show that each left-multiplication is a homomorphism from M to M. More precisely, for each $s \in R$, define $\varphi_s\colon M \to M$ by $\varphi_s(m) = sm$, and show that φ_s is a homomorphism.

(6.2.10) **Definitions.** Assume M is an R-module.

1) For $S \subseteq M$, the (unique) smallest submodule of M that contains S is called the ***submodule generated by S***, and is denoted $\langle S \rangle$.
2) If $\langle S \rangle = M$, then we say that S is a ***generating set*** of M.
3) M is ***cyclic*** if it is generated by one-element subset: $M = \langle m \rangle$ for some $m \in M$.
4) For $m \in M$, the ***annihilator*** of m in R is $\mathrm{Ann}_R(m) = \{ r \in R \mid rm = 0 \}$.
5) The ***annihilator*** of M in R is
$$\mathrm{Ann}_R(M) = \{r \in R \mid rm = 0 \text{ for all } m \in M\} = \{r \in R \mid rM = 0\} = \bigcap_{m \in M} \mathrm{Ann}_R(m).$$

(6.2.11) **Exercises.** Assume M and N are R-modules.

1) Show $\langle m\rangle = Rm$, for all $m \in M$.
2) More generally, show that if $S = \{s_1, s_2, \ldots, s_k\} \subseteq M$, then
$$\langle s_1, s_2, \ldots, s_k\rangle = Rs_1 + Rs_2 + \cdots + Rs_k.$$
3) Assume $\{s_1, s_2, s_3, \ldots, s_k\}$ is a generating set of M, and $m = s_1 + q_2s_2 + \cdots + q_ks_k$, for some $q_2, \ldots, q_k \in R$. Show $\{m, s_2, s_3, \ldots, s_k\}$ is also a generating set of M
4) Show $\mathrm{Ann}_R(m)$ is a left ideal in R, for all $m \in M$.
5) Show $\mathrm{Ann}_R(M)$ is a two-sided ideal in R.
6) For $M_1, M_2 \leq M$, show:
   (a) $\mathrm{Ann}_R(M_1) \cap \mathrm{Ann}_R(M_2) = \mathrm{Ann}_R(M_1 + M_2)$.
   (b) $\mathrm{Ann}_R(M_1) + \mathrm{Ann}_R(M_2) \subseteq \mathrm{Ann}_R(M_1 \cap M_2)$.
   (c) Equality does *not* always hold in (6b). Can you find an example?
7) Show that if I is any left ideal of R, then the module R/I is cyclic.
8) Let $m \in M$, so Rm is a cyclic submodule of M.
   (a) Show $\mathrm{Ann}_R(Rm) \subseteq \mathrm{Ann}_R(m)$.
   (b) Show the inclusion is an equality if R is commutative.
9) Show that if $M \cong N$, then $\mathrm{Ann}_R(M) = \mathrm{Ann}_R(N)$.

(6.2.12) **Definition.** The ***direct sum*** $M_1 \oplus M_2 \oplus \cdots \oplus M_k$ of R-modules $M_1, \ldots, M_k$ is the set $M_1 \times M_2 \times \cdots \times M_k$ with componentwise addition and scalar multiplication. More precisely, for $m_i, m_i' \in M_i$ and $r \in R$, we have

1) $(m_1, \ldots, m_k) + (m_1', \ldots, m_k') = (m_1 + m_1', \ldots, m_k + m_k')$, and
2) $r(m_1, \ldots, m_k) = (rm_1, \ldots, rm_k)$.

We use M^k as a shorthand for $M \oplus M \oplus \cdots \oplus M$, where there are k summands. In the special case where $M = R$, we often refer to R^k as a ***free*** module (cf. Exercise 6.4.7(1)).

(6.2.13) **Exercises.** Show:

1) If $M_i \cong N_i$ for $1 \leq i \leq k$, then $\bigoplus_{i=1}^k M_i \cong \bigoplus_{i=1}^k N_i$.
2) If $M_1, M_2 \leq M$, with $M_1 + M_2 = M$ and $M_1 \cap M_2 = \{0\}$, then $M \cong M_1 \oplus M_2$.
3) More generally, if $M_1, \ldots, M_k \leq M$, such that
   (a) $M_1 + \cdots + M_k = M$, and
   (b) for all i, we have $M_i \cap \sum_{j\neq i} M_j = \{0\}$,
   then $M \cong \bigoplus_{i=1}^k M_i$.
4) If N is a submodule of M, and there is a homomorphism $\varphi\colon M \to N$, such that $\varphi(n) = n$ for every $n \in N$, then $M \cong N \oplus \ker(\varphi)$.
5) If $m_1, \ldots, m_k \in M$, and we define $\varphi\colon R^k \to M$ by
$$\varphi(r_1, \ldots, r_k) = r_1m_1 + \cdots + r_km_k,$$
   then φ is a homomorphism.
6) If $M = \langle s_1, s_2, \ldots, s_k\rangle$, then $M \cong R^k/N$, for some submodule N of R^k.

(6.2.14) **Definition.** In the situation of Exercise 6.2.13(2), we say that M is the ***internal direct sum*** of M_1 and M_2.

§6.3. Isomorphism Theorems and the Correspondence Theorem

The three Isomorphism Theorems (1.4.10) and the Correspondence Theorem (1.4.12) of group theory have natural analogues in ring theory.

(6.3.1) **Notation.** In this section, M and N are always R-modules.

(6.3.2) **Proposition** (1st, 2nd, and 3rd Isomorphism Theorems for modules)**.**

1) *If φ is a homomorphism from M to N, then $M/\ker\varphi \cong \varphi(M)$. More precisely, an isomorphism $\overline{\varphi}\colon M/\ker\varphi \to \varphi(M)$ is obtained by defining $\overline{\varphi}(m + \ker\varphi) = \varphi(m)$.*
2) *Suppose $K, L \leq M$. Then $(K + L)/L \cong K/(K \cap L)$.*
3) *If $K \leq L \leq M$, then $L/K \leq M/K$, and $(M/K)/(L/K) \cong M/L$.*

(6.3.3) **Proposition** (Correspondence Theorem)**.** *Suppose N is a submodule of M. Then there is a one-to-one correspondence between the submodules of M that contain N and the submodules of M/N. Namely, the submodule of M/N corresponding to a submodule L of M is L/N.*

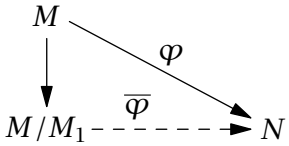
(6.3.4) **Definitions.**

1) M is ***simple*** if it is nontrivial and has no nontrivial, proper submodules.
2) A submodule N of M is ***maximal*** if N is proper, and is not contained in any larger, proper submodule of M. That is, if L is a submodule of M, such that $N \subseteq L \subseteq M$, then either $L = N$ or $L = M$.

(6.3.5) **Exercises.**

1) Show that if $M_1, M_2 \leq M$, with $M_1 + M_2 = M$, then $M/(M_1 \cap M_2) \cong M/M_1 \oplus M/M_2$.
2) Suppose $\varphi\colon M \to N$ is a homomorphism, and M_1 is a submodule of M that is contained in $\ker\varphi$. Show that a well-defined homomorphism $\overline{\varphi}\colon M/M_1 \to N$ can be obtained by defining $\overline{\varphi}(m + M_1) = \varphi(m)$.
3) Prove Propositions 6.3.2 and 6.3.3.
   [*Hint:* See the proofs of Propositions 1.4.10 and 1.4.12.]
4) Suppose N is a submodule of M. Show M/N is simple if and only if N is a maximal submodule of M.
5) Suppose M_1 and M_2 are submodules of M, such that M/M_1 is simple, and $M_2 \nsubseteq M_1$. Show $M_1 + M_2 = M$.
6) Suppose $M = Rm$ is cyclic. Show $M \cong R/\mathrm{Ann}_R(m)$.
7) Assume R is commutative, and let M and N be cyclic R-modules. Show $M \cong N$ if and only if $\mathrm{Ann}_R(M) = \mathrm{Ann}_R(N)$.
8) Suppose I is a two-sided ideal of R, and we would like to make M into an R/I module, by defining $(r + I)m = rm$. What would need to be true in order for this operation to be well-defined?

(6.3.6) *Remark.* To summarize the conclusion of Exercise 6.3.5(2), we may draw the following commutative diagram:



§6.4. Free modules and direct products

Recall that a subset of a vector space is a basis if it spans and is linearly independent. We make the same definitions for modules.

(6.4.1) **Notation.** As usual, M is an R-module.

(6.4.2) **Definitions.** Let $S = \{s_1, s_2, \ldots, s_k\}$ be a finite subset of M.

1) For any $r_1, \ldots, r_k \in R$, we call the expression $r_1 s_1 + \cdots + r_k s_k$ a ***linear combination*** of elements of S.
2) S ***spans*** M (or is a ***spanning set***) if every element of M is a linear combination of elements of S.

3) S is ***linearly independent*** if, for all $r_1, \dots, r_k \in R$, such that $r_1 s_1 + \cdots + r_k s_k = 0$, we have $r_1 = r_2 = \cdots = r_k = 0$.
4) S is a ***basis*** of M if it spans M and is linearly independent.

(6.4.3) **Exercise.** Let $S = \{s_1, s_2, \dots, s_k\}$ be a finite subset of M. Show that S is a basis of M if and only if every element can be written **uniquely** as a linear combination of elements of M. More precisely:

1) For all $m \in M$, there exist $r_1, \dots, r_k \in R$, such that $r_1 s_1 + \cdots + r_k s_k = m$.
2) If $r_1 s_1 + \cdots + r_k s_k = r_1' s_1' + \cdots + r_k' s_k'$, then $r_i = r_i'$ for all i.

(6.4.4) **Example.** $\{1\}$ is a basis of the regular R-module R.

It is a basic fact of linear algebra that every vector space has a basis. This is not always true for modules:

(6.4.5) **Definition.** M is ***free*** if it has a basis.

(6.4.6) **Exercises.** Let I be a nontrivial, proper two-sided ideal of R.

1) Show that the R-module R/I is not free.
[*Hint:* No nonempty subset is linearly independent.]
2) Recall that I is an R-module (see Example 6.1.4(4)). Show that if R is commutative, and the ideal I is not principal, then this module is not free.
[*Hint:* No two elements are linearly independent.]
3) Show that the $\mathbb{Z}$-module $\mathbb{Q}$ is not free.
[*Hint:* No two elements are linearly independent.]

In linear algebra, the cardinality of a basis is called the "dimension" of the vector space, and it is proved that any vector space of dimension n is isomorphic to F^n, where F is the field of scalars. Similar results are true in the theory of modules.

(6.4.7) **Exercises.** Show:

1) R^k is a free R-module (and has a basis of cardinality k).
2) Conversely, if M has a basis of cardinality k, then $M \cong R^k$.
3) If M and N both have a basis of cardinality k, then $M \cong N$.
4) If $S = \{s_1, s_2, \dots, s_k\}$ is a basis of M, and $n_1, n_2, \dots, n_k \in N$, then there is a unique homomorphism $\varphi\colon M \to N$, such that $\varphi(s_i) = n_i$ for each i.
5) If $N \leq M$, and $M/N \cong R^k$ is free, then $M \cong (M/N) \oplus N$.
[*Hint:* Use Exercise 6.2.13(4) to show that if $\{ s_1 + N, \dots, s_k + N \}$ is a basis of M/N, then $M = \langle s_1, \dots, s_k \rangle \oplus N$ (internal direct sum).]

We will sometimes want to be able to take the direct sum of infinitely many modules, not just finitely many. With Exercise 6.4.7(1) in mind, this should be done in such a way that if $M_i \cong R$ for all i, then $\bigoplus_{i=1}^{\infty} M_i$ is a free module, and a basis is given by $\{e_i\}$, where e_i has only one nonzero entry, which is in the ith spot. We begin by generalizing Definition 6.4.2 to allow bases that are infinite.

(6.4.8) **Definitions.** Let S be a (finite or infinite) subset of M.

1) For $k \in \mathbb{Z}^+$, $r_1, \dots, r_k \in R$, and $s_1, \dots, s_k \in S$, the expression $r_1 s_1 + \cdots + r_k s_k$ is called a ***linear combination*** of elements of S.
2) S ***spans*** M (or is a ***spanning set***) if every element of M is a linear combination of elements of S.
3) S is ***linearly independent*** if, for all $k \in \mathbb{Z}^+$, $r_1, \dots, r_k \in R$, and all pairwise distinct $s_1, \dots, s_k \in S$, such that $r_1 s_1 + \cdots + r_k s_k = 0$, we have $r_1 = r_2 = \cdots = r_k = 0$.
4) S is a ***basis*** of M if it spans M and is linearly independent.
5) M is ***free*** if it has a basis.

(6.4.9) **Exercise.** Let S be a (finite or infinite) subset of M. Show that S is a basis of M if and only if every element can be written **uniquely** as a linear combination of elements of M. More precisely:

1) For all $m \in M$, there exist $k \in \mathbb{Z}^+$, $r_1, \dots, r_k \in R$, and $s_1, \dots, s_k \in S$, such that $r_1 s_1 + \cdots + r_k s_k = m$.
2) If $r_1 s_1 + \cdots + r_k s_k = r'_1 s'_1 + \cdots + r'_\ell s'_\ell$, where
   - $r_1, r_2, \dots, r_k,\ r'_1, r'_2, \dots, r'_\ell \in R$,
   - $s_1, s_2, \dots, s_k$ are distinct elements of S, and
   - $s'_1, s'_2, \dots, s'_\ell$ are distinct elements of S,

   then $k = \ell$ and there exists $\sigma \in S_k$, such that $r_i = r'_{\sigma(i)}$ and $s_i = s'_{\sigma(i)}$ for all i.

The obvious generalization of Definition 6.2.12 leads to the following definition:

(6.4.10) **Definition.** The ***direct product*** $\prod_{i=1}^\infty M_i$ of a sequence of R-modules $M_1, M_2, \dots$ is the set

$$\{\, (m_1, m_2, \dots) \mid m_i \in M_i \,\}$$

with componentwise addition and scalar multiplication:

1) $(m_1, m_2 \dots) + (m'_1, m'_2 \dots) = (m_1 + m'_1, m_2 + m'_2, \dots)$, and
2) $r(m_1, m_2, \dots) = (rm_1, rm_2, \dots)$.

Unfortunately, however, this definition does not (usually) satisfy the natural analogue of Exercise 6.4.7(1):

(6.4.11) **Exercise.** Let $M_i = \mathbb{Z}$ for all i, let $M = \prod_{i=1}^\infty M_i$, and let $S = \{e_i\}_{i=1}^\infty$, where $e_i \in M$ is defined by $e_i(j) = \delta_i^j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$ Show that $\{e_i\}$ does not span the $\mathbb{Z}$-module M.

[*Hint:* Every linear combination of elements of S has only finitely many nonzero entries. Alternatively, only countably many elements of M are linear combinations of elements of S, but M is uncountable.]

For this reason (and others), the direct product defined in Definition 6.4.10 is often not nearly as useful as the following notion:

(6.4.12) **Definition.** The ***direct sum*** $\bigoplus_{i=1}^\infty M_i$ of a sequence of R-modules $M_1, M_2, \dots$ is the set

$$\{\, (m_1, m_2, \dots) \mid m_i \in M_i \text{ for all } i, \text{ and } m_i = 0 \text{ for all but finitely many } i \,\}$$

with componentwise addition and scalar multiplication.

(6.4.13) *Remark.* Definition 6.4.12 defines the direct sum of only countably many modules, but it is possible to take the direct sum of a larger collection of modules, as follows. Suppose we have a module M_i for each i in a set I (of arbitrary cardinality). (The set I is often called an "index set." In Definition 6.4.12, it is $\{1, 2, 3, \dots\}$.) Then

1) $\prod_{i \in I} M_i = \{\, f : I \to \bigcup_{i \in I} M_i \mid f(i) \in M_i \text{ for all } i \,\}$.
2) $\bigoplus_{i \in I} M_i = \{\, f \in \prod_{i \in I} M_i \mid f(i) = 0 \text{ for all but finitely many } i \,\}$.

(6.4.14) **Exercises.** Show:

1) $\bigoplus_{i=1}^\infty M_i$ is a submodule of $\prod_{i=1}^\infty M_i$.
2) The set $S = \{e_i\}$ defined in Exercise 6.4.11 is a basis of $\bigoplus_{i=1}^\infty R$.
3) For $i \in \mathbb{Z}^+$, let $\widehat{M}_i = \{\, (x_j) \in \bigoplus_{i=1}^\infty M_i \mid x_j = 0 \text{ for } j \neq i \,\}$, and define $\varphi_i : M_i \mapsto \widehat{M}_i$ by $\varphi(m) = (0, 0, \dots, 0, m, 0, 0, \dots)$. Then φ_i is an isomorphism.
4) (universality of the direct sum) Suppose we have a homomorphism $\varphi_i : M_i \to N$, for each $i = 1, 2, \dots$. Then:
   (a) for each $(m_1, m_2, \dots) \in \bigoplus_{i=1}^\infty M_i$, the sum $\sum_{i=1}^\infty \varphi_i(m_i)$ has only finitely many nonzero terms, so it specifies a well-defined element of N, and
   (b) we obtain a homomorphism $\varphi : \bigoplus_{i=1}^\infty M_i \to N$, by defining
   $$\varphi(m_1, m_2, \dots) = \textstyle\sum_{i=1}^\infty \varphi_i(m_i).$$
5) If
   (a) $\sum_{i=1}^\infty M_i = M$, and

(b) for all i, we have $M_i \cap \sum_{j \neq i} M_j = \{0\}$,

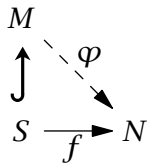
then $M \cong \bigoplus_{i=1}^{\infty} M_i$. (We say that M is the ***internal direct sum*** of $M_1, M_2, \ldots$.)

6) Conversely, suppose $M = \bigoplus_{i=1}^{\infty} M_i$. For $1 \leq i < \infty$, define

$$\widehat{M}_i = \{\, (m_1, m_2, \ldots) \in M \mid m_j = 0 \text{ for } j \neq i \,\}.$$

Then $\widehat{M}_i \cong M_i$ for all i, and $\bigoplus_{i=1}^{\infty} M_i$ is the internal direct sum of $\widehat{M}_1, \widehat{M}_2, \ldots$.

7) (compare with Theorem 4.2.9) M is free if and only if there is a subset S of M, such that every function f from S to any R-module N can be extended to a unique homomorphism φ from M to N.



8) Find three $\mathbb{Z}$-modules K, L, and M, such that $K \oplus M \cong L \oplus M$, but $K \ncong L$.

We have the following natural analogue of Corollary 4.2.13, which states that every group is a homomorphic image of a free group:

(6.4.15) **Exercise.** Show that every R-module is (isomorphic to) a quotient of a free R-module.

[*Hint:* See the proof of Corollary 4.2.12.]

(6.4.16) **Warning.** It is a basic fact of linear algebra that any two bases of a vector space have the same cardinality (which is called the "dimension" of the vector space). Unfortunately, the following example shows that this is not true for free modules over a general ring.

(6.4.17) **Exercise.** Let

- S be a ring,
- M be a free S-module that has a countably infinite basis $\{e_n\}_{i=1}^{\infty}$, and
- $R = \{\, \text{homomorphisms } \varphi \colon M \to M \,\}$.

Then R is a ring under addition and composition, and the regular R-module (obviously) has a basis with only one element. Show that it also has a basis with two elements.

[*Hint:* A basis $\{\varphi_0, \varphi_1\}$ is obtained by defining $\varphi_\epsilon(e_k) = \begin{cases} e_{(k+\epsilon)/2} & \text{if } k \equiv \epsilon \pmod 2, \\ 0 & \text{if } k \not\equiv \epsilon \pmod 2 \end{cases}$ for $\epsilon \in \{0, 1\}$.]

(6.4.18) **Exercise.** Show that if R is commutative, then it is true that any two bases of a free R-module always have the same cardinality (in contrast to the preceding example, in which R is not commutative).

[*Hint:* You may assume (without proof) that R has a maximal ideal I. (This fact will be proved in Exercise 7.4.6(2).) Let $F = R/I$, so F is a field. The cardinality of every basis of M is equal to the dimension of the F-vector space $M/(IM)$.]

Chapter 7

Modules over a PID

(7.0.1) **Assumption.** Assume R is a PID.

(7.0.2) **Example.** For any $\ell \in \mathbb{N}$ and $n_1, \ldots, n_k \in \mathbb{N}^+$, it is pretty obvious that the abelian group $\mathbb{Z}^\ell \times \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \cdots \times \mathbb{Z}_{n_k}$ is finitely generated.

A major theorem states that these are the only examples:

(7.0.3) **Theorem** (Fundamental Theorem of Finitely Generated Abelian Groups)**.**
Every finitely generated abelian group is a direct product of (*finitely many*) *cyclic groups.*

In other words, every finitely generated $\mathbb{Z}$-module is a direct sum of cyclic modules. In fact, this is true with any PID in the place of $\mathbb{Z}$:

(7.0.4) **Main Theorem** (Structure Theorem for Finitely Generated Modules over a PID)**.**
If R is a PID, then every finitely generated R-module is a direct sum of (*finitely many*) *cyclic modules.*

Proving this fundamental theorem is the main goal of this chapter.

§7.1. Torsion modules over a PID

(7.1.1) **Definition.**

- m is a ***torsion*** element of M if $rm = 0$, for some nonzero $r \in R$.
- M is a ***torsion*** module if every element of M is a torsion element.
- At the other extreme, M is ***torsion-free*** (or M has ***no torsion***) if 0 is the only torsion element of M.

(7.1.2) **Exercises.**

1) Show that if $\mathrm{Ann}_R(M) \neq \{0\}$, then M is a torsion module.
2) Prove that the converse holds if M is finitely generated.
3) Show the set M_{tors} of torsion elements of M is the (unique) largest torsion submodule of M. (In particular, M_{tors} is a submodule.)
4) Show M/M_{tors} has no torsion.
5) Show that if M is nonzero, torsion-free, and cyclic, then $M \cong R$ (so M is free).

We will begin the proof of Theorem 7.0.4 by looking at the case where M is a torsion module. (This is all that will be needed for our main applications.) The analogue in group theory would be to assume that G is finite. In that settting, it is important to study p-groups (especially Sylow p-subgroups). These are the groups in which the order of every element is a power of p. Here is the analogous notion in the theory of modules:

(7.1.3) **Definition.** Let M be an R-module, and let π be an irreducible element of R.

1) An element m of M is ***π-primary*** if there exists $n \in \mathbb{N}$, such that $\pi^n m = 0$.
2) M is ***π-primary*** if every element of M is π-primary.

(7.1.4) **Exercise.** Assume M is finitely generated. Show M is π-primary iff there exists $n \in \mathbb{N}$, such that $\pi^n M = 0$.

(7.1.5) **Exercise.** Show that quotients of primary modules are primary: if M is π-primary, and $N \leq M$, then M/N is π-primary.

3) The ***π-primary component*** M_π of M is the set of all π-primary elements of M.

(7.1.6) **Exercise.** Show that the π-primary component of M is the (unique) largest π-primary submodule of M. (In particular, it is a submodule of M.)

(7.1.7) **Exercise.** Suppose π and π' are irreducibles in R, and assume $M_\pi \neq \{0\}$. Show $M_\pi = M_{\pi'}$ iff π and π' are associates. (I.e., $\pi' = \pi u$ for some unit u.)

(7.1.8) **Proposition.** *Suppose*

- $m \in M$,
- $r_1 r_2 m = 0$, *for some* $r_1, r_2 \in R$, *and*
- $\gcd(r_1, r_2) = 1$.

Then there exist unique $m_1, m_2 \in M$, *such that* $m = m_1 + m_2$, $r_1 m_1 = 0$, *and* $r_2 m_2 = 0$.

Proof. Since R is a PID, there exist $a_1, a_2 \in R$, such that $a_1 r_1 + a_2 r_2 = \gcd(r_1, r_2) = 1$.

(Existence) Let $m_1 = a_2 r_2 m$ and $m_2 = a_1 r_1 m$. Then

$$m = (a_1 r_1 + a_2 r_2)m = a_1 r_1 m + a_2 r_2 m = m_2 + m_1 = m_1 + m_2.$$

Also, we have

$$r_1 m_1 = r_1(a_2 r_2 m) = a_2(r_1 r_2 m) = a_2 0 = 0$$

and

$$r_2 m_2 = r_2(a_1 r_1 m) = a_1(r_1 r_2 m) = a_1 0 = 0.$$

(Uniqueness) Suppose $m_1 + m_2 = m = m_1' + m_2'$. Let $y = m_1 - m_1' = m_2' - m_2$. Then

$$\begin{aligned} y &= (a_1 r_1 + a_2 r_2)y \\ &= a_1 r_1 (m_1 - m_1') + a_2 r_2 (m_2' - m_2) \\ &= a_1 r_1 m_1 - a_1 r_1 m_1' + a_2 r_2 m_2' - a_2 r_2 m_2 \\ &= a_1 0 - a_1 0 + a_2 0 - a_2 0 \\ &= 0, \end{aligned}$$

so $m_1 = m_1'$ and $m_2 = m_2'$. □

It is a special case of Theorem 3.2.4(7) that every finite abelian group is the direct product of its Sylow subgroups (because abelian groups are nilpotent). This fact generalizes to modules in the following way:

(7.1.9) **Corollary.** *Any torsion module over a PID is the direct sum of its primary components.*

Proof. If $rm = 0$ and the factorization of r into irreducibles is $r = \pi_1^{n_1} \pi_2^{n_2} \cdots \pi_k^{n_k}$, then Proposition 7.1.8 implies that m can be written uniquely in the form

$$m = m_1 + m_2 + \cdots + m_k,$$

where each m_i is π_i-primary. □

This reduces the proof of the Structure Theorem (7.0.4) for torsion modules to the special case of primary modules, which we will now prove. The proof uses the following definition:

(7.1.10) **Definition.** Let M be a finitely generated R-module. The ***rank*** of M is the smallest cardinality of a generating set of M. It is denoted $\operatorname{rank} M$.

(This means that if $\operatorname{rank} M = k$, then some generating set of M has k elements, but no generating set has less than k elements.)

(7.1.11) **Theorem.** *If M is a finitely generated, π-primary R-module, then M is a direct sum of cyclic submodules.*

Proof. Let $k = \operatorname{rank} M$. The proof is by induction on k, and the base case, where $\operatorname{rank} M = 1$, is trivial.

Define the ***valuation*** $v_\pi \colon R \to \mathbb{N} \cup \{\infty\}$ by

$$v_\pi(r) = \max\{\, \ell \in \mathbb{N} \mid \pi^\ell \mid r \,\}$$

(with $v_\pi(0) = \infty$). Choose a generating set $\{s_1, \ldots, s_k\}$ of M with cardinality k, and also choose $a_1, a_2, \ldots, a_k \in R$, such that $a_1 s_1 + \cdots + a_k s_k = 0$ and $a_1 \neq 0$. Among all the possible choices of $s_1, \ldots, s_k$ and $a_1, a_2, \ldots, a_k$, assume we have chosen one so that

$$v_\pi(a_1) \text{ is as small as possible.}$$

For convenience, let $\ell = v_\pi(a_1)$.

Since M is π-primary, there is some π^n, such that $\pi^n M = \{0\}$ (see Exercise 7.1.4). Write $a_1 = \pi^\ell a_1'$. Then $\gcd(\pi^n, a_1') = 1$, so there is some $b \in R$, such that $b a_1' \equiv 1 \pmod{\pi^n}$ (see Exercise 5.3.12(2b)). After multiplying $a_1, \ldots, a_k$ by b, we have $a_1 \equiv \pi^\ell \pmod{\pi^n}$, so there is no harm in assuming $a_1 = \pi^\ell$.

Since the elements of $\{s_1, \ldots, s_k\}$ can be permuted, the minimality of $\ell = v_\pi(a_1)$ implies that $\pi^\ell \mid a_i$ for all i. This means we may write $a_i = \pi^\ell q_i$. Let

$$s = s_1 + \sum_{i=2}^{k} q_i s_i \quad \text{and} \quad N = Rs_2 + \cdots + Rs_k.$$

Since $\operatorname{rank} N = k - 1 < \operatorname{rank} M$, we know, by the induction hypothesis, that N is a direct sum of cyclic modules.

To complete the proof, we show that $M = Rs \oplus N$. (Since Rs is a cyclic module, and N is a direct sum of cyclic modules, this implies that M is also a direct sum of cyclic modules.) It is easy to see that $Rs + N = M$ (see Exercise 6.2.11(3)).

So all that remains is to show that $Rs \cap N = \{0\}$. To this end, let $m \in Rs \cap N$. Then there exist $t_1, t_2, \ldots, t_k \in R$, such that $m = t_1 s$ and $m = t_2 s_2 + \cdots + t_k s_k$. Therefore

$$t_1 s - t_2 s_2 - t_3 s_3 - \cdots - t_k s_k = m - m = 0.$$

So the minimality of $\ell = v_\pi(a_1)$ implies $\pi^\ell \mid t_1$, which means $t_1 = q\pi^\ell$ for some $q \in R$. We also have

$$\pi^\ell s = a_1 \left(s_1 + \sum_{i=2}^{k} q_i s_i \right) = a_1 s_1 + \sum_{i=2}^{k} a_1 q_i s_i = \sum_{i=1}^{k} a_i s_i = 0.$$

Therefore $m = t_1 s = q(\pi^\ell s) = q(0) = 0$. Since m is an arbitrary element of $Rs \cap N$, we conclude that $Rs \cap N = \{0\}$, as desired. □

This establishes a refinement of the Structure Theorem (7.0.4) for the special case of torsion modules:

(7.1.12) **Corollary.** *Let M be a finitely generated torsion module over a PID R. Then M is a direct sum of finitely many cyclic modules $C_1, C_2, \ldots, C_k$, such that each C_i is primary.*

Proof. We know:

1) M is the direct sum of its (nonzero) primary components (see Corollary 7.1.9).

2) Each of these primary components is a direct sum of cyclic modules (by Theorem 7.1.11). Therefore, M is a direct sum of several submodules that are each a direct sum of primary cyclic modules. So M is the direct sum of all of those primary cyclic modules put together. □

The following example shows that the decomposition of M into a direct sum of cyclic modules is not usually unique (but see Remark 7.2.6).

(7.1.13) **Exercise.** Suppose $M = Rx \oplus Ry$ is the direct sum of two cyclic modules. Show that M is cyclic iff $\mathrm{Ann}_R(x) + \mathrm{Ann}_R(y) = R$.

§7.2. Completion of the proof of the Structure Theorem

(7.2.1) **Exercise.** Show that if $M \cong R^n$, then M has no torsion.

The key to completing the proof of the Structure Theorem (7.0.4) is to establish that the converse is true for finitely generated modules (over a PID):

(7.2.2) **Theorem.** *If M is a finitely generated, torsion-free R-module (and R is a PID), then $M \cong R^n$ for some $n \in \mathbb{N}$ (so M is free).*

(7.2.3) **Warning.** The assumption that M is finitely generated cannot be omitted from this theorem:

(7.2.4) **Exercise.** Show that $\mathbb{Q}$ is a torsion-free $\mathbb{Z}$-module (and $\mathbb{Z}$ is a PID), but $\mathbb{Q}$ is <u>not</u> a free $\mathbb{Z}$-module.

[*Hint:* $\mathbb{Q}$ is not cyclic, but every pair of nonzero submodules of $\mathbb{Q}$ has nonzero intersection.]

The proof of the theorem is postponed, but it is easy to derive the Structure Theorem from this:

(7.2.5) **Exercise.** Assume M is a finitely generated module over a PID R. Show that M is the direct sum of finitely many cyclic modules, each of which is either free or primary.

[*Hint:* Combine Exercise 6.4.7(5) with Theorem 7.2.2 and other results for a short, easy proof that $M \cong R^n \oplus M_{\mathrm{tors}}$, for some $n \in \mathbb{N}$. You may assume all of the exercises and other results stated previously, except the Structure Theorem for Modules over a PID.]

(7.2.6) *Remark.* The requirement that each of the cyclic modules is either free or primary makes the decomposition into cyclic modules unique, up to isomorphism and permutation of the summands (cf. Corollary 7.3.1).

In our applications, R is always a Euclidean domain, and there is a short proof in this case.

Proof of Theorem 7.2.2 in the special case where R is a Euclidean domain. This uses the same basic idea as the proof of Theorem 7.1.11.

Consider a generating set $S = \{s_1, \ldots, s_k\}$ of M with $k = \operatorname{rank} M$. We may assume S is not a basis (for otherwise M is free, so Exercise 6.4.7(2) tells us that $M \cong R^k$, so it is obvious that M is a direct sum of cyclic modules). Therefore, there exist $a_1, a_2, \ldots, a_k \in R$, not all 0, such that $a_1 s_1 + \cdots + a_k s_k = 0$. Among all such generating sets, and all such $a_1, a_2, \ldots, a_k$, with $a_1 \neq 0$, make a choice so that $d(a_1)$ is as small as possible.

We claim that $a_1 \mid a_i$ for all i. By symmetry, it suffices to show that $a_1 \mid a_2$. Since R is Euclidean, there exist $q, r \in R$, such that $a_2 = qa_1 + r$, and $d(r) < d(a_1)$. Let $m = s_1 + qs_2$. Then Exercise 6.2.11(3) tells us that $\{m, s_2, s_3, \ldots, s_k\}$ generates M. Also, we have

$$\begin{aligned} a_1 m + r s_2 + a_3 s_3 + \cdots + a_k s_k &= a_1(s_1 + q s_2) + + r s_2 + a_3 s_3 + \cdots + a_k s_k \\ &= a_1 s_1 + (a_1 q + r) s_2 + a_3 s_3 + \cdots + a_k s_k \\ &= a_1 s_1 + a_2 s_2 + a_3 s_3 + \cdots + a_k s_k \\ &= 0. \end{aligned}$$

Now, since $d(r) < d(a_1)$, the minimality of $d(a_1)$ implies that $r = 0$. From the definition of r, we conclude that $a_1 \mid a_2$, as desired.

From the claim, we know that we may write $a_i = a_1 q_i$ for each i. Let $s = s_1 + \sum_{i=2}^{k} q_i s_i$. Then

$$a_1 s = a_1 \left(s_1 + \sum_{i=2}^{k} q_i s_i \right) = a_1 s_1 + \sum_{i=2}^{k} a_1 q_i s_i = a_1 s_1 + \sum_{i=2}^{k} a_i s_i = 0.$$

Since M is torsion-free, we conclude that $s = 0$, so

$$s_1 = -\sum_{i=2}^{k} q_i s_i \in \langle s_2, s_3, \ldots, s_k \rangle.$$

Therefore

$$M = \langle s_1, s_2, s_3, \ldots, s_k \rangle = \langle s_2, s_3, \ldots, s_k \rangle.$$

This contradicts the fact that $\operatorname{rank} M = k$. □

To prepare for the proof for the general case where R is only assumed to be a PID (not a Euclidean domain), we develop some additional theory.

§7.2(i). Ascending chain condition for modules (optional)**.** We defined the ascending chain condition for rings (see Definition 5.3.8), but the notion also makes sense for modules:

(7.2.7) **Definition.** An R-module has the ***ascending chain condition*** (or ***acc***, for short) if every ascending chain of submodules is eventually constant. This is equivalent to any of the following three conditions:

1) There is no infinitely long *strictly* increasing chain of submodules.
2) Every submodule is finitely generated.
3) Every nonempty set of submodules has a maximal element.

(*Alternate terminology:* Just as for rings, saying that a module is ***Noetherian*** is the same as saying that it has the acc.)

(7.2.8) **Exercises.** Assume R is **commutative**.

1) Show that R has the acc (as a ring) if and only if the regular R-module has the acc (as an R-module).
2) Suppose M is an R-module with the acc.
   (a) Show that every proper submodule of M is contained in a maximal submodule of M.
   (b) Show that every cyclic submodule of M is contained in a maximal cyclic submodule of M.
3) Suppose N is a submodule of the R-module M.
   (a) Show that if M has the acc, then N and M/N both have the acc.
   (b) (slightly harder) Conversely, show that if N and M/N both have the acc, then M has the acc.
   (c) Show that if M_1 and M_2 have the acc, then $M_1 \oplus M_2$ has the acc.
4) Show that if R is a PID, then every finitely generated R-module has the acc.
   [*Hint:* Exercise 6.2.13(6).]

§7.2(ii). Torsion-free modules (optional)**.** This section proves the Structure Theorem (7.0.4) under the assumption that M is torsion-free. The proof is by induction on $\operatorname{rank} M$, and the base case is easy (see Exercise 7.1.2(5)). The induction step will use a lemma.

(7.2.9) **Lemma.** *If M is torsion-free, and $\langle m \rangle$ is a maximal cyclic submodule, then $M/\langle m \rangle$ is torsion-free.*

Proof. Let $\overline{}$ be the natural homomorphism from M to $M/\langle m\rangle$.

Suppose $\overline{x}$ is any torsion element of $\overline{M}$. This means there is some nonzero $r \in R$, such that $r\overline{x} = \overline{0}$, so $rx \in \langle m\rangle$. Thus, we may write $rx = sm$ for some $s \in R$.

Let $d = \gcd(r, s)$, so we may write $r = r'd$ and $s = s'd$. Then

$$d(r'x - s'm) = dr'x - ds'm = rx - sm = 0.$$

Since M is torsion-free, this implies $r'x - s'm = 0$. Now, $\gcd(r', s') = 1$, so there exist $u, v \in R$, such that $ur' + vs' = 1$. Then

$$m = (ur' + vs')m = ur'm + vs'm = ur'm + vr'x = r'(um + vx) \in \langle um + vx\rangle.$$

Combined with the maximality of $\langle m\rangle$, this implies that $\langle m\rangle = \langle um + vx\rangle$, so r' must be a unit. (Why?) Therefore

$$x = (r')^{-1}r'x = (r')^{-1}s'm \in \langle m\rangle,$$

so $\overline{x} = 0$. Since $\overline{x}$ is an arbitrary torsion element of $\overline{M}$, this implies that $\overline{M}$ is torsion-free. □

(7.2.10) **Proposition.** *If R is a PID and M is a finitely generated, torsion-free R-module, then M is free, so $M \cong R^n$ for some $n \in \mathbb{N}$.*

Proof. The proof is by induction on $\operatorname{rank} M$ (see Definition 7.1.10). The base case, where $\operatorname{rank} M = 1$, is (see Exercise 7.1.2(5)).

Let $k = \operatorname{rank} M$, choose a generating set S of M with $\#S = k$, and let $s \in S$. Since M has the acc, we may assume $\langle s\rangle$ is a maximal cyclic submodule, so Lemma 7.2.9 tells us that $M/\langle s\rangle$ is torsion-free. Furthermore, $S \smallsetminus \{s\}$ generates $M/\langle s\rangle$, so $k(M/\langle s\rangle) = k - 1$. Therefore, by the induction hypothesis, we know that $M/\langle s\rangle$ is free. So $M/\langle s\rangle \cong R^{k-1}$.

Now, from Exercises 6.4.7(5) and 7.1.2(5), we see that

$$M \cong (M/\langle s\rangle) \oplus \langle s\rangle \cong R^{k-1} \oplus R \cong R^k,$$

so M is free. □

§7.3. Fundamental Theorem of Finitely Generated Abelian Groups

Abelian groups are just another name for $\mathbb{Z}$-modules (see Example 6.1.4(2)) and $\mathbb{Z}$ is a PID (in fact, Example 5.3.3(1) tells us that it is Euclidean), so the Fundamental Theorem of Finitely Generated Abelian Groups (7.0.3) is a special case of Theorem 7.0.4. And we can say a bit more if we consider only the abelian groups that are finite:

(7.3.1) **Corollary** (Fundamental Theorem of Finite Abelian Groups)**.** *Every finite, abelian group is a direct product of cyclic groups of prime-power order. Furthermore, the nontrivial factors in the product are unique up to a permutation.*

(7.3.2) **Examples.**

1) $\mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_4 \cong \mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_3$, because the factors in each product are the same, but in a different order.
2) $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_4 \ncong \mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_4$, because $\mathbb{Z}_2$ occurs three times as a factor on the left-hand side, but only once on the right-hand side (or, if you prefer, because $\mathbb{Z}_4$ occurs only once as a factor on the left-hand side, but twice on the right-hand side).

Proof of Corollary 7.3.1. Let G be a finite, abelian group, and let P be a Sylow p-subgroup of G, for some prime p that divides $|G|$. We know that every abelian group is a $\mathbb{Z}$-module (see Example 6.1.4(2)), so G and P are $\mathbb{Z}$-modules.

Since P is finite, it is obviously finitely generated, so the Fundamental Theorem of Finitely Generated Abelian Groups (7.0.3) tells us that P is a direct sum of cyclic submodules. In group-theoretic terminology, this means that P is a direct product of cyclic subgroups. Since P is a p-group, the order of every subgroup of P is a power of p. Therefore, we see that P is a direct product of cyclic groups of prime-power order.

We now know that every Sylow subgroup of G is a direct product of cyclic groups of prime-power order. We also know that G is the direct product of its Sylow subgroups (from either

Corollary 7.1.9 or Theorem 3.2.4(7), whichever you prefer). Putting these together tells us that G is the direct product of cyclic groups of prime-power order.

We leave the uniqueness of the decomposition as an exercise (see Exercise 7.3.3(2)). □

(7.3.3) **Exercises.** Let C_t be the cyclic group of order t, for each $t \in \mathbb{N}$.

1) Suppose $n_1, n_2, \ldots, n_k \in \mathbb{N}$. For every $m \in \mathbb{N}^+$, show
$$\frac{|\{\, x \in C_{p^{n_1}} \times \cdots \times C_{p^{n_k}} \mid x^{p^m} = 1 \,\}|}{|\{\, x \in C_{p^{n_1}} \times \cdots \times C_{p^{n_k}} \mid x^{p^{m-1}} = 1 \,\}|} = p^r, \text{ where } r = |\{\, i \mid n_i \geq m \,\}|.$$
2) Show that if $C_{p^{n_1}} \times \cdots \times C_{p^{n_k}} \cong C_{p^{m_1}} \times \cdots \times C_{p^{m_\ell}}$ with $n_1 \geq n_2 \geq \cdots \geq n_k \geq 1$ and $m_1 \geq m_2 \geq \cdots m_\ell \geq 1$, then $k = \ell$ and $n_i = m_i$ for every i.

Corollary 7.3.1 makes it easy to find all the abelian groups of any given order. For example, we will find all of the abelian groups of order p^5 in Example 7.3.5 below, but let us first make an important observation:

(7.3.4) **Observation.** Suppose G is an abelian group of order p^n, for some prime p and some $n \in \mathbb{N}$. From Corollary 7.3.1, we know G can be written (uniquely) in the form $\mathbb{Z}_{p^{n_1}} \times \mathbb{Z}_{p^{n_2}} \times \cdots \mathbb{Z}_{p^{n_k}}$, with $n_1 \geq n_2 \geq \cdots \geq n_k \geq 1$. For this direct product to be of the desired order, we must have $n_1 + n_2 + \cdots + n_k = n$.

By definition, a ***partition*** of a natural number n is a representation of n as an unordered sum of positive integers. (By "unordered," we mean that two representations are considered the same if they are obtained by changing the order of the summands. For example, $4 + 2 + 2$ and $2 + 4 + 2$ represent the same partition of 8, because they consist of the same terms in a different order. It is traditional to list the summands in decreasing order.) Thus, we have a one-to-one correspondence between the abelian groups of order p^n and the partitions of n.

(7.3.5) **Example.** We can find all of the abelian groups of order p^5, for any prime p. The partitions of 5 are:
$$5,\ 4+1,\ 3+2,\ 3+1+1,\ 2+2+1,\ 2+1+1+1,\ 1+1+1+1+1.$$
So (up to isomorphism) the abelian groups of order p^5 are:
$$\mathbb{Z}_{p^5}, \qquad \mathbb{Z}_{p^4} \times \mathbb{Z}_{p^1}, \qquad \mathbb{Z}_{p^3} \times \mathbb{Z}_{p^2}, \qquad \mathbb{Z}_{p^3} \times \mathbb{Z}_p \times \mathbb{Z}_p,$$
$$\mathbb{Z}_{p^2} \times \mathbb{Z}_{p^2} \times \mathbb{Z}_p, \qquad \mathbb{Z}_{p^2} \times \mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}_p, \qquad \mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}_p.$$

The above technique can be used to find all of the abelian groups of any prime-power order. For orders that are not a power of a prime, we first find all of the possible Sylow p-subgroups, for each p dividing the order, and then take the direct products of these Sylow subgroups.

(7.3.6) **Example.** Find (up to isomorphism) all of the abelian groups of order 12,250.

Solution. We have $12{,}250 = 2^1 \cdot 5^3 \cdot 7^2$.

- The only partition of 1 is 1, so the Sylow 2-subgroup must be $\mathbb{Z}_2$.
- The partitions of 3 are 3, 2 + 1, and 1 + 1 + 1, so the Sylow 5-subgroup can be any of the following:
$$\mathbb{Z}_{5^3} = \mathbb{Z}_{125}, \quad \mathbb{Z}_{5^2} \times \mathbb{Z}_5 = \mathbb{Z}_{25} \times \mathbb{Z}_5, \quad \mathbb{Z}_5 \times \mathbb{Z}_5 \times \mathbb{Z}_5.$$
- The partitions of 2 are 2 and 1 + 1, so the Sylow 7-subgroup can be either of the following:
$$\mathbb{Z}_{7^2} = \mathbb{Z}_{49}, \quad \mathbb{Z}_7 \times \mathbb{Z}_7.$$

Therefore, the abelian groups of order 12,250 are:
$$\begin{array}{ccc} \mathbb{Z}_2 \times \mathbb{Z}_{125} \times \mathbb{Z}_{49}, & \mathbb{Z}_2 \times \mathbb{Z}_{25} \times \mathbb{Z}_5 \times \mathbb{Z}_{49}, & \mathbb{Z}_2 \times \mathbb{Z}_5 \times \mathbb{Z}_5 \times \mathbb{Z}_5 \times \mathbb{Z}_{49}, \\ \mathbb{Z}_2 \times \mathbb{Z}_{125} \times \mathbb{Z}_7 \times \mathbb{Z}_7, & \mathbb{Z}_2 \times \mathbb{Z}_{25} \times \mathbb{Z}_5 \times \mathbb{Z}_7 \times \mathbb{Z}_7, & \mathbb{Z}_2 \times \mathbb{Z}_5 \times \mathbb{Z}_5 \times \mathbb{Z}_5 \times \mathbb{Z}_7 \times \mathbb{Z}_7. \end{array}$$
□

(7.3.7) **Exercises.**

1) Find (up to isomorphism) all of the abelian groups of each of the following orders:
   (a) 36
   (b) 40
   (c) $5^2 \cdot 11^4$
   (d) $3^2 \cdot 7^3 \cdot 19^2$
   (e) $210 = 2^1 \cdot 3^1 \cdot 5^1 \cdot 7^1$
2) Show that every nontrivial, finite, abelian group can be written uniquely as a direct product $\mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \cdots \times \mathbb{Z}_{n_k}$, where $n_{i+1} \mid n_i$ for all i (and $n_i > 1$).
3) Suppose K, L, and M are finite abelian groups, such that $K \oplus M \cong L \oplus M$. Show that $K \cong L$.

(7.3.8) *Remark.* The finiteness assumption cannot be omitted in Exercise 7.3.7(3) (see Exercise 6.4.14(8)). However, it suffices to assume that the groups are finitely generated, rather than being finite.

§7.4. Zorn's Lemma (advanced)

We have seen that if the group G is finite, then every proper subgroup of G is contained in a maximal subgroup (see Lemma 3.2.13). Similarly, it is immediate from Definition 5.3.8(4) that if R is a commutative ring with the acc, then every proper ideal of R is contained in a maximal ideal. We will see that the conclusion is true even without assuming that R has the acc, but the proof is more sophisticated (cf. Example 7.4.4).

In this section, we explain a very useful principal (called "Zorn's Lemma") that establishes the existence of maximal objects in a wide variety of situations.

(7.4.1) **Definitions.**

1) A ***partial order*** on a set $\mathcal{P}$ is a binary relation $\prec$ on $\mathcal{P}$ that is transitive and antireflexive. More precisely, for $x, y, z \in \mathcal{P}$, we have:
   - (transitive) If $x \prec y$ and $y \prec z$, then $x \prec z$.
   - (antireflexive) $x \not\prec x$.
2) A ***partially ordered set*** is a set $\mathcal{P}$ together with a partial order $\prec$ on $\mathcal{P}$.
3) A subset $\mathcal{C}$ of a partially ordered set $\mathcal{P}$ is a ***chain*** (or is ***totally ordered***) if, for all $x, y \in \mathcal{C}$, either $x \prec y$ or $y \prec x$ or $x = y$.
4) Let $S \subseteq \mathcal{P}$. An element b of $\mathcal{P}$ is an ***upper bound*** of S of $\mathcal{P}$ if $s \preceq b$, for all $s \in S$.
5) An element m of $\mathcal{P}$ is ***maximal*** if there does not exist $x \in \mathcal{P}$, such that $m \prec x$.

(7.4.2) **Theorem** ("Zorn's Lemma"). *If $(\mathcal{P}, \prec)$ is a partially ordered set, such that every chain in $\mathcal{P}$ has an upper bound, then $\mathcal{P}$ has a maximal element.*

Idea of proof. Suppose there is no maximal element. (This will lead to a contradiction.) For any $x_n \in \mathcal{P}$, we know, by assumption, that x_n is not maximal, so there exists $x_{n+1} \succ x_n$. Therefore, by induction, we may construct a sequence $\{x_n\}_{n=1}^{\infty}$, such that $x_1 \prec x_2 \prec x_3 \prec \cdots$. Then $\{x_n\}$ is a chain in $\mathcal{P}$, so, by assumption, it has an upper bound x_1'.

Now, by repeating the argument, we can construct an infinite sequence $\{x_n'\}_{n=1}^{\infty}$, such that $x_1' \prec x_2' \prec x_3' \prec \cdots$. Then $\{x_n'\}$ is a chain in $\mathcal{P}$, so, by assumption, it has an upper bound x_1''.

Repeating the argument over and over, we construct more and more distinct elements of $\mathcal{P}$. If $\mathcal{P}$ is countable, then a contradiction can be obtained by repeating this construction an uncountable number of times. In general, a contradiction can be obtained by repeating the construction a number of times that is greater than the cardinality of $\mathcal{P}$ (using a process called "transfinite induction").

See (7.4.8) below for a proof that does not require transfinite induction (but is less intuitive). □

The empty set is (vacuously) a chain in $\prec$, so the assumption that every chain has an upper bound implies that $\mathcal{P} \neq \emptyset$. The empty chain is often an annoyance in proofs, so it may be a bit easier to apply the following reformulated version.

(7.4.3) **Theorem** ("Zorn's Lemma")**.** *If $(\mathcal{P}, \prec)$ is a nonempty partially ordered set, such that every nonempty chain in $\mathcal{P}$ has an upper bound, then $\mathcal{P}$ has a maximal element.*

Here is a sample application:

(7.4.4) **Example.** Every nonzero commutative ring R has a maximal ideal.

Proof. Let $\mathcal{P}$ be the set of all proper ideals of R. (Note that $\mathcal{P} \neq \emptyset$, because $\{0\}$ is a proper ideal of R.) Then $\mathcal{P}$ is partially ordered by inclusion. (That is, the relation $\subset$ is a partial order on $\mathcal{P}$.)

We claim that every nonempty chain in $\mathcal{P}$ has an upper bound. To see this, let $\mathscr{C}$ be any nonempty chain in $\mathcal{P}$, and let $J = \bigcup_{I \in \mathscr{C}} I$. It is not difficult to see that J is a proper ideal:

- For $r \in R$ and $j \in J$, the definition of union implies there exists $I_0 \in \mathscr{C}$, such that $j \in I_0$. Then $rj \in rI_0 \subseteq I_0 \subseteq \bigcup_{I \in \mathscr{C}} I = J$. So J is closed under scalar multiplication.
- For $j_1, j_2 \in J$, the definition of union implies there exists $I_1, I_2 \in \mathscr{C}$, such that $j_1 \in I_1$ and $j_2 \in I_2$. Since $\mathscr{C}$ is totally ordered by inclusion, either $I_1 \subseteq I_2$ or $I_2 \subseteq I_1$. Assume, without loss of generality, that $I_1 \subseteq I_2$. Then
$$j_1 + j_2 \in I_1 + I_2 \subseteq I_2 + I_2 = I_2 \subseteq \bigcup_{I \in \mathscr{C}} I = J,$$
so J is closed under addition.
- Since $\mathscr{C}$ is nonempty, there is some $I_0 \in \mathscr{C}$. Then $0 \in I_0 \subseteq \bigcup_{I \in \mathscr{C}} I = J$, so $J \neq \emptyset$.
- For each $I \in \mathcal{P}$ we have $1 \notin I$ (since $\mathcal{P}$ is the set of proper ideals). Therefore $1 \notin \bigcup_{I \in \mathscr{C}} I = J$, so the ideal J is proper.

So $J \in \mathcal{P}$. Also, by the definition of J, we have $I \subseteq J$ for all $I \in \mathscr{C}$. Therefore J is an upper bound of $\mathscr{C}$.

Hence, Zorn's Lemma implies that $\mathcal{P}$ has a maximal element M. By the definition of $\mathcal{P}$, this means that M is a maximal ideal of R. □

(7.4.5) *Remark.* Example 7.4.4 is a typical application of Zorn's Lemma in algebra. Namely, we often encounter a situation in which:

1) We want to find a maximal element in a collection of subsets of some set X, where "maximal" means that it is not a proper subset of any other member of the collection.
2) It is not difficult to see that the union of any totally ordered collection of the objects is again either in our collection or is the entire set X.
3) We have an obstruction (the element 1 in the above example) to show that the union of any totally ordered collection of the objects is not all of X.

(7.4.6) **Exercises.**

1) Show that every nonzero ring (not necessarily commutative) has
   (a) a maximal left ideal,
   (b) a maximal right ideal, and
   (c) a maximal two-sided ideal.
2) Show that every vector space has a basis.
3) Suppose g is a nontrivial element of a group G, and let $\mathcal{H}$ be the set of all subgroups of G that do **not** contain g. The set $\mathcal{H}$ is partially ordered under inclusion. Show it has a maximal element.
4) Suppose N is a submodule of the R-module M. Show there is a submodule N' of M, such that
   (a) $N \cap N' = \{0\}$, and
   (b) if N'' is any submodule of M that properly contains N', then $N \cap N'' \neq \{0\}$.

5) Let $G = \mathbb{Q}$ be the additive group of rational numbers.
   (a) Show that G has no maximal subgroups.
   (b) Show that the collection $\mathcal{P}$ of proper subgroups of G is partially ordered by inclusion, and that the union of any totally ordered collection of subgroups of G is a subgroup of G.
   (c) Find a chain $\mathscr{C}$ in $\mathcal{P}$, such that $\bigcup_{H\in\mathscr{C}} H = G$.

   This illustrates that Part (3) of Remark 7.4.5 is crucial for applications of Zorn's Lemma.
6) Show that if a binary relation $\prec$ is transitive and antireflexive, then it is antisymmetric: if $x \prec y$, then $y \not\prec x$.

(7.4.7) *Remark.* Zorn's Lemma is equivalent to the famous ***Axiom of Choice***:

Suppose
- *X and Y are sets,*
- *$\mathcal{P}(Y)$ is the collection of all subsets of Y,*
- *$A\colon X \to \mathcal{P}(Y)$, and*
- *$A(x) \neq \emptyset$, for all $x \in X$.*

Then there exists $a\colon X \to Y$, such that $a(x) \in A(x)$, for all $x \in X$.

So anything that can be proved by using Zorn's Lemma can also be proved by using the Axiom of Choice. However, Zorn's Lemma is often easier to apply for problems that typically arise in algebra.

(7.4.8) **Proof of Zorn's Lemma** (optional)**.** By assumption, each chain $\mathscr{C}$ has an upper bound $b_0(\mathscr{C})$ in $\mathcal{P}$. Now suppose $\mathcal{P}$ does not have a maximal element. (This will lead to a contradiction.) Then, for each chain $\mathscr{C}$, there exists some $b(\mathscr{C}) \succ b_0(\mathscr{C})$ (because $b_0(\mathscr{C})$ is not maximal). Note that, for every chain $\mathscr{C}$:
- $b(\mathscr{C})$ is an upper bound for $\mathscr{C}$, and
- $b(\mathscr{C}) \notin \mathscr{C}$.

Definitions. Let $\mathscr{C}$ be a chain.
1) $\mathscr{C}$ is ***well-ordered*** if every nonempty subset of $\mathscr{C}$ has a smallest element.
2) $\mathscr{C}^{\prec c} = \{\, a \in \mathscr{C} \mid a \prec c \,\}$ for any $c \in \mathscr{C}$. We call $\mathscr{C}^{\prec c}$ an ***initial segment*** of $\mathscr{C}$.
3) $\mathscr{C}$ is ***generated by b*** if $b(\mathscr{C}^{\prec c}) = c$ for all $c \in \mathscr{C}$.
4) $\mathcal{B}$ is the collection of all well-ordered chains in $\mathcal{P}$ that are generated by b.
5) $\mathcal{B}^* = \bigcup_{\mathscr{C}\in\mathcal{B}} \mathscr{C}$ is the union of all the chains in $\mathcal{B}$.

Step 1. If $\mathscr{C}_1, \mathscr{C}_2 \in \mathcal{B}$, then either $\mathscr{C}_1$ is an initial segment of $\mathscr{C}_2$, or $\mathscr{C}_2$ is an initial segment of $\mathscr{C}_1$. We may assume $\mathscr{C}_1 \neq \mathscr{C}_2$. Then, since $\mathscr{C}_1$ and $\mathscr{C}_2$ are well-ordered, the symmetric difference $\mathscr{C}_1 \triangle \mathscr{C}_2$ has a minimal element c. We may assume that $c \in \mathscr{C}_1 \smallsetminus \mathscr{C}_2$ (by interchanging $\mathscr{C}_1$ and $\mathscr{C}_2$ if necessary). We may also assume $\{\, x \in \mathscr{C}_2 \mid x \succeq c \,\}$ is nonempty (otherwise $\mathscr{C}_2$ is an initial segment of $\mathscr{C}_1$), so it has a minimal element m. Then

$$\begin{aligned} c &= b(\mathscr{C}_1^{\prec c}) && (c \in \mathscr{C}_1 \text{ and } \mathscr{C}_1 \text{ is generated by } b)\\ &= b(\mathscr{C}_2^{\prec c}) && (\text{minimality of } c \text{ implies } \mathscr{C}_1^{\prec c} = \mathscr{C}_2^{\prec c})\\ &= b(\mathscr{C}_2^{\prec m}) && (\mathscr{C}_2^{\prec c} = \mathscr{C}_2^{\prec m} \text{ by the minimality of } m)\\ &= m && (m \in \mathscr{C}_2 \text{ and } \mathscr{C}_2 \text{ is generated by } b)\\ &\in \mathscr{C}_2. \end{aligned}$$

This contradicts the fact that $c \in \mathscr{C}_1 \smallsetminus \mathscr{C}_2$ (so $c \notin \mathscr{C}_2$).

Step 2. $\mathcal{B}^*$ is a well-ordered chain that is generated by b (so $\mathcal{B}^* \in \mathcal{B}$). Given any $x, y \in \mathcal{B}^*$, there exist $\mathscr{C}_x, \mathscr{C}_y \in \mathcal{B}$ with $x \in \mathscr{C}_x$ and $y \in \mathscr{C}_y$. From Step 1, we may assume that $\mathscr{C}_x$ is an initial segment of $\mathscr{C}_y$, so $x \in \mathscr{C}_x \subseteq \mathscr{C}_y$. Therefore, the (totally ordered) chain $\mathscr{C}_y$ contains both x and y, so either $x \preceq y$ or $y \preceq x$. So $\mathcal{B}^*$ is a chain.

Now, let X be any nonempty subset of $\mathcal{B}^*$. Choose some $x \in X$, and a chain $\mathscr{C}_x \in \mathcal{B}$ that contains x. From Step 1, we see that $\mathscr{C}_x$ is an initial segment of $\mathcal{B}^*$. So the minimal element of $\mathscr{C}_x \cap X$ is the minimal element of X. Therefore, $\mathcal{B}^*$ is well-ordered.

We now show that $\mathcal{B}^*$ is generated by b. Given $c \in \mathcal{B}^*$, there is some $\mathscr{C} \in \mathcal{B}$, such that $c \in \mathscr{C}$. From Step 1, we see that $\mathscr{C}$ is an initial segment of $\mathcal{B}^*$, so $(\mathcal{B}^*)^{\prec c} = \mathscr{C}^{\prec c}$. Therefore $b\big((\mathcal{B}^*)^{\prec c}\big) = b(\mathscr{C}^{\prec c}) = c$, since $\mathscr{C}$ is generated by b.

Step 3. We have a contradiction. Let $\mathscr{C}^+ = \mathcal{B}^* \cup b(\mathcal{B}^*)$. Then, since $\mathcal{B}^*$ is a well-ordered chain that is generated by b, it is easy to see that $\mathscr{C}^+$ is also a well-ordered chain that is generated by b. This means $\mathscr{C}^+ \in \mathcal{B}$, so $\mathscr{C}^+ \subseteq \bigcup_{\mathscr{C} \in \mathcal{B}} \mathscr{C} = \mathcal{B}^*$. But this is clearly not true, because $b(\mathcal{B}^*) \in \mathscr{C}^+$, but $b(\mathcal{B}^*) \notin \mathcal{B}^*$. □

Part III

Linear Algebra

Chapter 8

Review

Linear Algebra is the study of vector spaces (and related topics). We will begin our discussion by quickly reviewing some of the basic material that is in a typical undergraduate course, but from a more advanced viewpoint that assumes familiarity with the theory of rings and modules. Unlike an undergraduate course, where the scalars are assumed to be real numbers (or perhaps complex numbers), we will often allow them to come from any field F.

(8.0.1) **Terminology.**

- If F is a field, then F-modules (usually denoted V or W) are called ***vector spaces*** over F.
- Homomorphisms between vector spaces are often called ***linear transformations***.
  - A homomorphism from V to itself may be called a ***linear operator*** on V.
- Submodules of V are usually called ***subspaces*** of V. (They may also be called "linear subspaces" or "vector subspaces.")

(8.0.2) **Notation.** Throughout this chapter:

- F is a field.
- U, V, W, and Z are vector spaces over F.

§8.1. Basis, dimension, coordinates, etc.

(8.1.1) **Definitions** (cf. Definition 6.4.2)**.** Let $S = \{v_1, v_2, \ldots, v_k\}$ be a finite subset of V.

1) For any $\alpha_1, \ldots, \alpha_k \in R$, we call the expression $\alpha_1 v_1 + \cdots + \alpha_k v_k$ a ***linear combination*** of elements of S.
2) S ***spans*** V if every element of V is a linear combination of elements of S.
3) S is ***linearly independent*** if, for all $\alpha_1, \ldots, \alpha_k \in F$, such that $\alpha_1 v_1 + \cdots + \alpha_k v_k = 0$, we have $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$.
4) S is a ***basis*** of V if it spans V and is linearly independent.
5) V is ***finite-dimensional*** if it has a spanning set that is finite.

(8.1.2) **Exercise.** Suppose $S = \{v_1, v_2, \ldots, v_k\}$ is a subset of V that is **not** linearly independent. Show there exists $v \in S$, such that v is a linear combination of elements of $S \smallsetminus \{v\}$.

We will be mostly interested in vector spaces that are finite-dimensional.

(8.1.3) **Example.** The ***standard basis*** of $\mathbb{R}^n$ is $\{e_1, \ldots, e_n\}$, where every coordinate of e_i is 0, except that the ith coordinate is 1. For example, the standard basis of $\mathbb{R}^4$ is

$$\{(1,0,0,0),\ (0,1,0,0),\ (0,0,1,0),\ (0,0,0,1)\}.$$

(8.1.4) **Proposition.** *Assume V is finite-dimensional.*

1) *V has a basis. In fact,*
   (a) *every spanning set of vectors contains a basis, and*
   (b) *every linearly independent set of vectors is contained in a basis.*
2) *All bases of V have the same cardinality.*

   Definition. This cardinality is called the ***dimension*** of V, and is denoted $\dim V$.
3) *A finite subset $S = \{v_1, v_2, \ldots, v_k\}$ is a basis of V if and only if every element can be written **uniquely** as a linear combination of elements of S. More precisely:*
   (a) *For all $v \in V$, there exist $\alpha_1, \ldots, \alpha_k \in F$, such that $\alpha_1 v_1 + \cdots + \alpha_k v_k = v$.*
   (b) *If $\alpha_1 v_1 + \cdots + \alpha_k v_k = \alpha'_1 v_1 + \cdots + \alpha'_k v_k$, then $\alpha_i = \alpha'_i$ for all i.*
4) *If W is any proper subspace of V, then $\dim W < \dim V$.*
5) (Rank-Nullity Theorem) *If $T\colon V \to W$ is a linear transformation, then*
$$\dim T(V) = \dim V - \dim \ker T.$$
6) *If $T\colon V \to V$ is a linear operator, then the following are equivalent:*
   (a) *T is invertible.* (b) *T is one-to-one.* (c) *T is onto.*
7) *For $n \in \mathbb{Z}^+$, we have $\dim V = n \Leftrightarrow V \cong F^n$. More precisely, if $\mathcal{B} = \{v_1, \ldots, v_n\}$ is any ordered basis of V, then:*
   - *An isomorphism $T\colon F^n \to V$ can be defined by*
$$T(\alpha_1, \ldots, \alpha_n) = \alpha_1 v_1 + \cdots + \alpha_n v_n.$$
   - *Conversely,*
$$\alpha_1 v_1 + \cdots + \alpha_n v_n \mapsto (\alpha_1, \ldots, \alpha_n)$$
   *is an isomorphism from V to F^n. When $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$, we call $(\alpha_1, \ldots, \alpha_n)$ the **coordinates** of v (with respect to the basis $\mathcal{B}$).*

   (8.1.5) **Notation.** Whenever it is convenient, we will identify F^n with the space of column vectors of length n, by identifying $(\alpha_1, \ldots, \alpha_n)$ with the matrix at right. When $(\alpha_1, \ldots, \alpha_n)$ are the coordinates of a vector v, we denote this matrix by $[v]_{\mathcal{B}}$.
$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$$
8) *The coordinates of v usually depend on the choice of the basis $\mathcal{B}$. More precisely, if C is another basis, then there exists a (unique) **change-of-basis matrix** $P_C^{\mathcal{B}}$, such that*
$$[v]_C = P_C^{\mathcal{B}} [v]_{\mathcal{B}} \text{ for all } v \in V.$$
   Namely, $P_C^{\mathcal{B}} = \begin{bmatrix} [v_1]_C & \cdots & [v_n]_C \end{bmatrix}$ is the matrix whose ith column is $[v_i]_C$. This is an $n \times n$ matrix. Note that the uniqueness implies $(P_C^{\mathcal{B}})^{-1} = P_{\mathcal{B}}^C$. (In particular, a change-of-basis matrix is always invertible.)

(8.1.6) **Exercises.** Assume $\dim V = n$, and let S be a set of precisely n elements of V. Show:

1) If S is linearly independent, then S is a basis of V.
2) If S spans V, then S is a basis of V.

(8.1.7) **Definition.** For $n \in \mathbb{Z}^+$, we use $\mathrm{Mat}_{n\times n}(F)$ to denote the set of all $n \times n$ matrices with entries in F. It is a ring (under the usual addition and multiplication of matrices), and is also a vector space over F (of dimension n^2).

(8.1.8) **Terminology.** Since $\mathrm{Mat}_{n\times n}(F)$ is both a ring and a vector space over F, it is said to be an ***algebra*** over F, but we do not need this terminology.

Every linear transformation can be represented by a matrix:

(8.1.9) **Proposition.** *Let $\mathcal{B} = \{v_1, \dots, v_n\}$ be a basis of the vector space V (so $\dim V = n$), and let $T\colon V \to V$ be a linear operator.*

1) *There is a unique matrix $[T]_{\mathcal{B}} \in \mathrm{Mat}_{n\times n}(F)$, such that $[Tv]_{\mathcal{B}} = [T]_{\mathcal{B}}[v]_{\mathcal{B}}$ for every $v \in V$. Namely,*
$$[T]_{\mathcal{B}} = \begin{bmatrix} [Tv_1]_{\mathcal{B}} & [Tv_2]_{\mathcal{B}} & \cdots & [Tv_n]_{\mathcal{B}} \end{bmatrix}$$
is the matrix whose ith column is the coordinate vector of Tv_i.
2) *The map $T \mapsto [T]_{\mathcal{B}}$ is an F-linear isomorphism from the ring of linear transformations of V to $\mathrm{Mat}_{n\times n}(F)$.*
3) *The matrix $[T]_{\mathcal{B}}$ usually depends on the choice of the basis $\mathcal{B}$. More precisely, if C is another basis of V, and $P_C^{\mathcal{B}}$ is the corresponding change-of-basis matrix, then*
$$[T]_C = P_C^{\mathcal{B}}\,[T]_{\mathcal{B}}\,(P_C^{\mathcal{B}})^{-1}.$$
*Therefore, similar matrices can be thought of as being representations of the same linear transformation with respect to different bases. (Recall that matrices $A, B \in \mathrm{Mat}_{n\times n}(F)$ are **similar** if there is an invertible matrix $P \in \mathrm{Mat}_{n\times n}(F)$, such that $PAP^{-1} = B$. This is an equivalence relation.)*

§8.2. Determinants, eigenvalues, and eigenvectors

(8.2.1) **Assumption.** In this section, V is always finite-dimensional.

Recall that the ***determinant*** $\det A$ of a square matrix A is an element of F.

(8.2.2) **Definition.** The ***determinant*** $\det A$ of an $n \times n$ matrix $A = [a_{i,j}]$ is defined inductively as follows:

- For $n = 1$, we have $\det[a_{1,1}] = a_{1,1}$.
- If $n > 1$, then, for any fixed i (with $1 \le i \le j$), we have
$$\det A = \sum_{j=1}^{n} (-1)^{i+j}\, a_{i,j}\, \det A^{i,j},$$
where $A^{i,j}$ is the $(n-1)\times(n-1)$ matrix that is obtained by deleting the ith row and jth column of A. (This formula is called "expanding along the ith row.)

(8.2.3) **Example.** By expanding along the first row, we have
$$\det\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = a \det\begin{bmatrix} e & f \\ h & i \end{bmatrix} - b\begin{bmatrix} d & f \\ g & i \end{bmatrix} + c\begin{bmatrix} d & e \\ g & h \end{bmatrix} = a(ei - fh) - b(di - fg) + c(dh - eg).$$

(8.2.4) **Proposition.** *Assume $A, B \in \mathrm{Mat}_{n\times n}(F)$.*

1) $\det I = 1$, *where I is the $n \times n$ identity matrix.*
2) *If $A = (a_{i,j})$ is **upper-triangular** (that is, if $a_{i,j} = 0$ whenever $i > j$), then $\det A = \prod_{i=1}^{n} a_{i,i}$ is the product of the diagonal entries of A.*
3) *If B is obtained from A by multiplying one of the rows (or columns) of A by a scalar α, then $\det B = \alpha \det A$.*
4) *If B is obtained from A by interchanging two of the rows (or two of the columns) of A, then $\det B = -\det A$.*
5) *If B is obtained from A by adding a scalar multiple of one of the rows of A to some other row of A, then $\det B = \det A$. (And the same is true with columns in the place of rows.)*
6) $\det(AB) = (\det A)(\det B)$.
7) *Similar matrices have the same determinant: if P is invertible, then $\det(PAP^{-1}) = \det A$.*
8) *A is invertible if and only if $\det A \neq 0$.*

(8.2.5) **Definition.** For $A \in \mathrm{Mat}_{m\times m}$ and $B \in \mathrm{Mat}_{n\times n}$, we use $A \oplus B$ to denote the matrix

$$\left[\begin{array}{c|c} A & 0_{m\times n} \\ \hline 0_{n\times m} & B \end{array}\right] \in \mathrm{Mat}_{(m+n)\times(m+n)}(F).$$

More generally, if $A_1, \ldots, A_n$ are square, then

$$A_1 \oplus A_2 \oplus \cdots \oplus A_n = \begin{bmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_n \end{bmatrix}.$$

A matrix of this form (with $n > 1$) is called "***block-diagonal***."

(8.2.6) **Proposition.** *The determinant of a block-diagonal matrix is the product of the determinants of its blocks:*

$$\det(A_1 \oplus A_2 \oplus \cdots \oplus A_n) = (\det A_1)(\det A_2) \cdots (\det A_n).$$

Similar matrices always have the same determinant (see Proposition 8.2.4(7)), so the following definition is independent of the choice of the basis $\mathcal{B}$.

(8.2.7) **Definition.** Assume $\dim V < \infty$. The ***determinant*** of a linear operator $T\colon V \to V$ is defined to be $\det[T]_{\mathcal{B}}$, for any basis $\mathcal{B}$ of V.

(8.2.8) **Definition.** Let $T\colon V \to V$ be a linear operator.

1) A vector $v \in V$ is an ***eigenvector*** of T if $v \neq 0$ and there exists $\lambda \in F$, such that $Tv = \lambda v$. The scalar λ is called an ***eigenvalue*** of T.
2) For each eigenvalue λ, the set $V_\lambda = \{\, v \in V \mid Tv = \lambda v \,\}$ is the corresponding ***eigenspace***. This is a subspace of V (because it is the kernel of the linear transformation $T - \lambda I$).
3) The ***characteristic polynomial*** of T is the polynomial $p(x) = \det(xI - T)$, where, as usual, I is the identity map.

Note that the characteristic polynomial is a monic polynomial whose degree is $\dim V$. (A polynomial is ***monic*** if its leading coefficient is 1. This means $p(x) = \sum_{i=0}^{n} a_i x^i$, with $a_n = 1$.)

The following important fact is a consequence of the observation that λ is an eigenvalue if and only if the operator $\lambda I - T$ has a kernel (and is therefore not invertible):

(8.2.9) **Exercise.** Show that the eigenvalues of a linear operator are the roots of its characteristic polynomial that are in F. More precisely, suppose $\lambda \in F$ and $p(x)$ is the characteristic polynomial of a linear transformation $T\colon V \to V$ (and V is finite-dimensional). Then λ is an eigenvalue of T if and only if $p(\lambda) = 0$.

(8.2.10) **Corollary.** *The eigenvalues of an upper-triangular matrix are its diagonal entries.*

Proof. If $A = (a_{i,j})$ is upper-triangular, then $xI - A$ is also upper triangular, with diagonal entries $x - a_{1,1}, x - a_{2,2}, \ldots, x - a_{n,n}$. Then, since the determinant of any upper-triangular matrix is the product of its diagonal entries (see Proposition 8.2.4(2)), we have

$$\det(xI - A) = (x - a_{1,1})(x - a_{2,2}) \cdots (x - a_{n,n}),$$

and the roots of this characteristic polynomial are obviously $a_{1,1}, a_{2,2}, \ldots, a_{n,n}$, the diagonal entries of A. □

(8.2.11) **Definitions.**

1) A matrix is ***triangularizable*** if it is similar to an upper-triangular matrix.
2) A linear operator T on a vector space V is ***triangularizable*** if there is a basis $\mathcal{B}$ of V, such that the matrix $[T]_{\mathcal{B}}$ is upper-triangular.

Corollary 8.2.10 implies that if a matrix (or linear operator) is triangularizable, then all the roots of its characteristic polynomial must be in F. The converse is true:

(8.2.12) **Proposition.** *Let $p(x)$ be the characteristic polynomial of a linear operator $T\colon V \to V$. A linear operator T on a vector space V is **triangularizable** if and only if $p(x)$ factors into a product of linear polynomials.*

Proof. We need only prove ($\Leftarrow$), since the other direction follows from Corollary 8.2.10. The proof is by induction on $\dim V$. The base case is easy, since every 1×1 matrix is upper triangular.

Assume, now, that $n = \dim V > 1$. By assumption, we may write

$$p(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n),$$

with each $\lambda_i \in F$. Let

- v_1 be an eigenvector corresponding to the eigenvalue λ_1, so $Tv_1 = \lambda_1 v_1$, and
- $\overline{V} = V/\langle v_1 \rangle$, so $\dim \overline{V} = n - 1$.

Since $T(\langle v_1 \rangle) \subseteq \langle v_1 \rangle$, T induces a well-defined linear operator $\overline{T}\colon \overline{V} \to \overline{V}$.

Choose $v_2, \ldots, v_n \in V$, such that $\overline{v}_2, \ldots, \overline{v}_n$ is a basis $\overline{\mathcal{B}}$ of $\overline{V}$. (For the moment, any basis will do, but, later in the proof, we will choose a basis more carefully.) Letting $\mathcal{B} = \{v_1, \ldots, v_n\}$, and recalling that $Tv_1 = \lambda_1 v_1$, we see that

$$[T]_{\mathcal{B}} = \left[\begin{array}{c|ccc} \lambda_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & [\overline{T}]_{\overline{\mathcal{B}}} & \\ 0 & & & \end{array}\right], \quad \text{so} \quad [xI - T]_{\mathcal{B}} = \left[\begin{array}{c|ccc} x - \lambda_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & [xI - \overline{T}]_{\overline{\mathcal{B}}} & \\ 0 & & & \end{array}\right] \tag{8.2.13}$$

By expanding the determinant along the first column, this block-diagonal form implies that

$$\det(xI - T) = (x - \lambda_1) \cdot \det(xI - \overline{T}),$$

so the characteristic polynomial of $\overline{T}$ is $p(x)/(x - \lambda_1) = (x - \lambda_2) \cdots (x - \lambda_n)$, which is a product of linear factors.

Hence, by induction on $\dim V$, we may assume that the basis $\overline{\mathcal{B}}$ has been chosen so that $[\overline{T}]_{\overline{\mathcal{B}}}$ is upper triangular. Then (8.2.13) shows that $[T]_{\mathcal{B}}$ is upper triangular. □

We can combine Proposition 8.2.12 with the following important fact (which means that $\mathbb{C}$ is "algebraically closed"):

(8.2.14) **Theorem** (Fundamental Theorem of Algebra)**.** *Every nonconstant polynomial in $\mathbb{C}[x]$ has a root in $\mathbb{C}$.*

Idea of proof. We describe one of the many different proofs of this theorem. Others can be found in textbooks on Complex Analysis or Field Theory, or on wikipedia:

`https://en.wikipedia.org/wiki/Fundamental_theorem_of_algebra#Proofs`

Suppose $f(x) \in \mathbb{C}[x]$, and $f(x)$ has no roots. Assume, for simplicity, that $f(x)$ is monic, and let $n = \deg f(x)$. For $r \geq 0$ and $t \in [0, 2\pi]$, let $p_r(t) = f(re^{it})$. For each fixed r, the function p_r is a closed path in $\mathbb{C} \smallsetminus \{0\}$, and therefore has a winding number $w(p_r)$, which is the net number of times that p_r winds around the origin (counter-clockwise).

The path p_0 is constant, so $w(p_0) = 0$. Since $w(p_r)$ is always an integer, but is also a continuous function of r, this implies that $w(p_r) = 0$ for all r. However, when r is large, the leading term x^n

of $f(x)$ is much larger than all others, so it is clear that $w(p_r) = n$ in this case. We conclude that $n = 0$, so $f(x)$ is constant. □

(8.2.15) **Corollary.** *Every linear operator on any vector space over $\mathbb{C}$ is triangularizable.*

Proof. The Fundamental Theorem of Algebra (8.2.14) implies that every nonconstant, irreducible polynomial in $\mathbb{C}[x]$ is linear, so every nonconstant polynomial in $\mathbb{C}[x]$ is a product of linear factors. In particular, If T is a linear operator on a vector space V over $\mathbb{C}$, then the characteristic polynomial of T is a product of linear factors. So Proposition 8.2.12 tells us that T is triangularizable. □

(8.2.16) **Example.** Let $A = \left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right]$, so the characteristic polynomial of A is $x^2 + 1$. Since the polynomial $x^2 + 1$ has no real roots, we see that A is <u>not</u> triangularizable over $\mathbb{R}$. That is, there does not exist an invertible matrix $P \in \mathrm{Mat}_{2\times 2}(\mathbb{R})$, such that PAP^{-1} is upper-triangular. In other words, if $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ is defined by $Tv = Av$, then there is no basis $\mathcal{B}$ of $\mathbb{R}^2$, such $[T]_{\mathcal{B}}$ is upper triangular.

On the other hand, we know from Corollary 8.2.15 that A must be similar to an upper-triangular matrix over $\mathbb{C}$. Indeed, for $P = \left[\begin{smallmatrix} 1 & i \\ i & 1 \end{smallmatrix}\right]$, we have $P^{-1}AP = \left[\begin{smallmatrix} i & 0 \\ 0 & -i \end{smallmatrix}\right]$, which is diagonal, not just upper triangular.

(8.2.17) **Exercise.** Show that eigenvectors corresponding to different eigenvalues are linearly independent.

More precisely, assume

1) $T\colon V \to V$ is a linear operator,
2) $v_1, \dots, v_k$ are eigenvectors of T,
3) λ_i is the eigenvector corresponding to v_i (i.e., $Tv_i = \lambda_i v_i$), and
4) $\lambda_i \neq \lambda_j$ for $i \neq j$.

Show that $v_1, \dots, v_k$ are linearly independent.

[*Hint:* Consider a linear combination that is 0, and has the minimum possible number of nonzero coefficients.]

§8.3. Diagonalizability

Diagonal matrices are easy to work with, but, unfortunately, not every matrix is similar to a diagonal matrix (even when its characteristic polynomial is a product of linear factors):

(8.3.1) **Exercise.** Show that the matrix $\left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$ is not similar to any diagonal matrix, even though its characteristic polynomial x^2 is a product of linear factors.

[*Hint:* If D is a diagonal matrix, with $d_1, \dots, d_n$ along the main diagonal, then D^2 has $d_1^2, \dots, d_n^2$ along the main diagonal. Hence, if $D \neq 0$, then $D^2 \neq 0$.]

(8.3.2) **Exercise.** For a linear operator $T\colon V \to V$ (with $\dim V < \infty$), show that the following are equivalent:

1) T is diagonalizable (that is, there is a basis $\mathcal{B}$ of V, such that $[T]_{\mathcal{B}}$ is a diagonal matrix).
2) There is a basis of V that consists entirely of eigenvectors of T.
3) V is spanned by the eigenvectors of T.

(8.3.3) **Exercise.** Suppose $T\colon V \to V$ is a linear operator, with $\dim V = n < \infty$. Show that if T has n distinct eigenvalues, then it is diagonalizable.

(8.3.4) **Definition.** For any polynomial $p(x) = \sum_{i=0}^{r} a_i x^i \in F[x]$ and $A \in \mathrm{Mat}_{n\times n}(F)$, we define

$$p(A) = \sum_{i=0}^{r} a_i A^i \in \mathrm{Mat}_{n\times n}(F).$$

For any fixed A, the map $p(x) \mapsto p(A)$ is a ring homomorphism.

(8.3.5) **Proposition.** *For any $A \in \mathrm{Mat}_{n\times n}(F)$, there is a unique monic polynomial $m(x) \in F[x]$, such that, for all $p(x) \in F[x]$, we have*

$$p(A) = 0 \iff m(x) \mid p(x).$$

*We call $m(x)$ the **minimal polynomial** of A.*

Proof. Let $I = \{\, p(x) \in F[x] \mid p(A) = 0 \,\}$. It is easy to verify that I is an ideal in the ring $F[x]$. (In fact, it is the kernel of the homomorphism $p(x) \mapsto p(A)$.) Also, since $\mathrm{Mat}_{n\times n}(F)$ is n^2-dimensional, the powers $I, A, A^2, A^3, \ldots, A^{n^2}$ cannot be linearly independent, so there exist $a_0, \ldots, a_{n^2} \in F$, not all 0, such that $\sum_{i=0}^{n^2} a_i A^i = 0$. In other words, if we let $p_0(x) = \sum_{i=0}^{n^2} a_i x^i$, then $p_0(x) \in I$, so the ideal I is nonzero. Since $F[x]$ is a PID (in fact, it is Euclidean), the ideal must be generated by some nonzero $m(x) \in F[x]$. This means that $p(x) \in I \iff m(x) \mid p(x)$.

The generator of I is unique up to multiplication by a scalar (that is, by a unit in the ring $F[x]$), so there is a unique choice of $m(x)$ that is monic. □

(8.3.6) *Remark.* The proof shows that $\deg m(x) \leq n^2$. In fact, the famous Cayley-Hamilton Theorem (10.2.17) will show that $\deg m(x) \leq n$.

(8.3.7) **Proposition.** *Let $A \in \mathrm{Mat}_{n\times n}(F)$. Then A is diagonalizable* (*via some invertible matrix $P \in \mathrm{Mat}_{n\times n}(F)$*) *if and only if the minimal polynomial of A is a product of linear factors in $F[x]$ and has no repeated roots.*

Proof. Let $m(x)$ be the minimal polynomial of A.

($\Rightarrow$) Suppose $A = \mathrm{diag}(a_1, \ldots, a_n)$ is a diagonal matrix with $a_1, \ldots, a_n$ on the main diagonal. Let

- $\lambda_1, \ldots, \lambda_k$ be its distinct eigenvalues (so $\{a_1, \ldots, a_n\} = \{\lambda_1, \ldots, \lambda_k\}$), and
- $p(x) = \prod_{j=1}^{k} (x - \lambda_j)$, so $p(a_i) = 0$ for all i.

Then, since A is diagonal, we have

$$p(A) = p\big(\mathrm{diag}(a_1, \ldots, a_n)\big) = \mathrm{diag}\big(p(a_1), p(a_2), \ldots, p(a_n)\big) = (0, 0, \ldots, 0) = 0.$$

Therefore $m(x) \mid p(x)$. From its definition, it is obvious that that $p(x)$ is a product of linear factors in $F[x]$, and has no repeated roots. The same must be true of the divisor $m(x)$.

($\Leftarrow$) By assumption, we may write $m(x) = \prod_{i=1}^{k} (x - \lambda_i)$, where $\lambda_1, \ldots, \lambda_k$ are distinct. For each i, let

$$p_i(x) = \prod_{j \neq i} (x - \lambda_i) = \frac{m(x)}{x - \lambda_i}.$$

Since $\gcd\big(p_1(x), p_2(x), \ldots, p_k(x)\big) = 1$, and $F[x]$ is a PID, there exist polynomials $u_i(x) \in F[x]$, such that

$$u_1(x)\, p_1(x) + u_2(x)\, p_2(x) + \cdots u_k(x)\, p_k(x) = 1.$$

Let $v \in F^n$ be arbitrary. For $1 \leq i \leq k$, let $v_i = u_i(A)\, p_i(A)\, v$. Then

$$(A - \lambda_i) v_i = u_i(A)\, \big((A - \lambda_i)\, p_i(A)\big) v = u_i(A)\, m(A)\, v = u_i(A) \cdot 0 \cdot v = 0,$$

so v_i is an eigenvector of A (with eigenvalue λ_i).

Furthermore, we have

$$v_1 + \cdots v_k = \sum_{i=1}^{k} u_i(A)\, p_i(A)\, v = \left(\sum_{i=1}^{k} u_i(A)\, p_i(A) \right) v = (I) v = v.$$

Since v is arbitrary, and $v_1, v_2, \ldots, v_k$ are eigenvectors, this implies that the eigenvectors of A span V. So Exercise 8.3.2 implies that A is similar to a diagonal matrix. □

(8.3.8) **Exercise.** A square matrix A is said to be ***nilpotent*** if there is some $k \in \mathbb{N}$, such that $A^k = 0$. For each n, show that 0 is the only $n \times n$ nilpotent matrix that is diagonalizable.

[*Hint:* This is a generalization of Exercise 8.3.1.]

(8.3.9) *Remark.* Similar matrices have the same minimal polynomial (why?). Therefore, any linear transformation $T\colon V \to V$ has a well-defined minimal polynomial (if $\dim V < \infty$), and Proposition 8.3.7 applies to it.

(8.3.10) **Exercise.** Assume T is a linear operator on a finite-dimensional vector space V.

1) Show that if $T^2 = T$, then T is diagonalizable.
2) Assume W is a subspace of V that is ***T*-invariant**. (This means $T(W) \subseteq W$, so the restriction $T|_W$ is a linear operator on W.) Show that if T is diagonalizable, then $T|_W$ is also diagonalizable.

Chapter 9

Bilinear forms and Hermitian forms

(9.0.1) **Notation.** Assume F is a field. (Usually, it will be $\mathbb{R}$ or $\mathbb{C}$.)

§9.1. Real symmetric matrices are diagonalizable (optional)

(9.1.1) **Definitions.**

1) Recall that the ***transpose*** of an $m \times n$ matrix $A = (a_{i,j})$ is the $n \times m$ matrix $A^\mathsf{T} = (a_{j,i})$ whose columns are the rows of A. For example,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}^\mathsf{T} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}^\mathsf{T} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}.$$

2) A matrix $A = (a_{i,j})$ is ***symmetric*** if it is equal to its transpose: $A^\mathsf{T} = A$. (This means that $a_{i,j} = a_{j,i}$ for all i, j, so A is symmetric across the main diagonal.) Every symmetric matrix must have the same number of rows as columns: it is a square matrix.

(9.1.2) **Examples.**

1) Every diagonal matrix is symmetric. (In particular, every 1×1 matrix is symmetric.)
2) Symmetric matrices that are 2×2 or 3×3 are of the following forms: $\begin{bmatrix} a & b \\ b & c \end{bmatrix}, \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}$.
3) An upper-triangular matrix cannot be symmetric unless it is also lower-triangular, and is therefore a diagonal matrix.

(9.1.3) **Exercises.**

1) Show that transpose is a linear operator on $\mathrm{Mat}_{m\times n}(F)$: $(sA + tB)^\mathsf{T} = sA^\mathsf{T} + tB^\mathsf{T}$ for $s, t \in F$ and $A, B \in \mathrm{Mat}_{m\times n}(F)$.
2) Show that the set of $n \times n$ symmetric matrices is a vector subspace of $\mathrm{Mat}_{n\times n}(F)$.
3) Show that the transpose of a product is the product of the transposes *in the reverse order:* if $A \in \mathrm{Mat}_{m\times n}(F)$ and $B \in \mathrm{Mat}_{n\times p}(F)$, then $(AB)^\mathsf{T} = B^\mathsf{T} A^\mathsf{T}$.
4) Show that the set of $n \times n$ symmetric matrices is closed under powers: if A is symmetric and $k \in \mathbb{N}$, then A^k is symmetric.
5) Show that the set of $n \times n$ symmetric matrices is **not** closed under multiplication (if $n \geq 2$).
6) Show $(A^\mathsf{T})^\mathsf{T} = A$ for all $A \in \mathrm{Mat}_{m\times p}(F)$.
7) Let $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$. Show that if there is an orthogonal matrix $P \in \mathrm{Mat}_{n\times n}(\mathbb{R})$, such that PAP^{-1} is a diagonal matrix, then A is symmetric. (Recall that P is ***orthogonal*** if $P^\mathsf{T} = P^{-1}$.)

(9.1.4) *Remark.* The converse of Exercise 9.1.3(7) is true (see Remark 9.1.13).

Although some matrices are not diagonalizable (see, for example, Exercise 8.3.1), we will see that every symmetric matrix with real entries is diagonalizable:

(9.1.5) **Theorem.** *Let $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$. If A is symmetric, then A is diagonalizable.*

More precisely, A is diagonalizable over $\mathbb{R}$: there is an invertible matrix $P \in \mathrm{Mat}_{n\times n}(\mathbb{R})$, such that PAP^{-1} is a diagonal matrix. The remainder of this section provides a proof of this important theorem, by using the following relation between transposes and the dot product:

(9.1.6) **Exercise.** For $v, w \in \mathbb{R}^n$ and $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$, show that

$$v \cdot Aw = (A^{\mathsf{T}}v) \cdot w,$$

where the ***dot product*** on $\mathbb{R}^n$ is defined by

$$(v_1, v_2, \ldots, v_n) \cdot (w_1, w_2, \ldots, w_n) = v_1w_1 + v_2w_2 + \cdots + v_nw_n.$$

[*Hint:* Recalling that we identify $(v_1, v_2, \ldots, v_n)$ with the corresponding column vector (see Notation 8.1.5), and identifying $\mathrm{Mat}_{1\times 1}(\mathbb{R})$ with $\mathbb{R}$ in the obvious way, the dot product may be written in terms of matrix multiplication as: $v \cdot w = v^{\mathsf{T}} w$.]

This has the following consequences:

(9.1.7) **Exercise.** Let A be a symmetric matrix in $\mathrm{Mat}_{n\times n}(\mathbb{R})$.

1) For $v \in \mathbb{R}^n$, such that $A^2v = 0$, show $Av = 0$.
   [*Hint:* If $A^2v = 0$, then $v \cdot A^2v = 0$.]
2) Show that if A is nilpotent, then $A = 0$.
3) Show that the minimal polynomial of A has no repeated real roots: if $m(x)$ is the minimal polynomial of A, and $c \in \mathbb{R}$, then $m(x)$ is not divisible by $(x - c)^2$.
   [*Hint:* $A - cI$ is symmetric. (*Why?*)]

To show that a real symmetric matrix A is diagonalizable, we wish to show that its minimal polynomial is a product of linear factors in $\mathbb{R}[x]$ and has no repeated roots (see Proposition 8.3.7). Exercise 9.1.7(3) deals with the question of repeated roots, but we still need to show that the minimal polynomial is a product of linear factors. For this, we extend the dot product on $\mathbb{R}^n$ to a certain product on $\mathbb{C}^n$:

(9.1.8) **Definition.** The (standard) ***Hermitan form*** on $\mathbb{C}^n$ is the function $\mathbb{C} \times \mathbb{C} \to \mathbb{C}$ that is defined by

$$\big\langle (z_1, z_2, \ldots, z_n) \,\big|\, (w_1, w_2, \ldots, w_n) \big\rangle = z_1\overline{w_1} + z_2\overline{w_2} + \cdots + z_n\overline{w_n}.$$

(Recall that the ***conjugate*** of a complex number $z = x + iy$ is the complex number $\overline{z} = x - iy$.) This is like the dot product, except that the second factor in each summand has been conjugated.

The following crucial property is a good reason for including complex conjugation in this definition:

(9.1.9) **Exercise.** For $v \in \mathbb{C}^n$, show that if $\langle v \mid v \rangle = 0$, then $v = 0$.

Corresponding to Exercise 9.1.6, we have the following observation:

(9.1.10) **Exercise.** Suppose $v, w \in \mathbb{C}^n$ and $A \in \mathrm{Mat}_{n\times n}(\mathbb{C})$. Show that $\langle v \mid Aw \rangle = \langle A^*v \mid w \rangle$, where A^* is the ***adjoint*** (or ***conjugate-transpose***) of $A = (a_{i,j})$, defined by

$$A^* = \overline{A^{\mathsf{T}}} = \overline{A}^{\mathsf{T}} = (\overline{a_{j,i}}).$$

[*Hint:* Note that $\langle v \mid w \rangle = w^* v$.]

(9.1.11) **Lemma.** *Every eigenvalue of a real symmetric matrix is real.*

Proof. Assume $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ is symmetric. If $Av = \lambda v$ (with $v \neq 0$ and $\lambda \in \mathbb{C}$), then

$$\begin{aligned}
\langle (A-\overline{\lambda}I)v \mid (A-\overline{\lambda}I)v\rangle &= \langle (A-\overline{\lambda}I)^*\,(A-\overline{\lambda}I)v \mid v\rangle && \text{(Exercise 9.1.10)}\\
&= \langle (A^*-\lambda I)\,(A-\overline{\lambda}I)v \mid v\rangle && ((sA+tB)^* = \overline{s}\,A^* + \overline{t}\,B^* \text{ for } s,t\in\mathbb{C})\\
&= \langle (A-\lambda I)\,(A-\overline{\lambda}I)v \mid v\rangle && \begin{pmatrix} A \text{ is symmetric, so } A^{\mathsf{T}} = A,\\ \text{and } A \text{ is a real matrix, so } \overline{A} = A\end{pmatrix}\\
&= \langle (A-\overline{\lambda}I)(A-\lambda I)v \mid v\rangle && \begin{pmatrix} A \text{ commutes with itself, and}\\ \text{the scalar matrices } \overline{\lambda}I \text{ and } \lambda I\\ \text{commute with everything}\end{pmatrix}\\
&= \langle (A-\overline{\lambda}I)\,0 \mid v\rangle && (Av = \lambda v)\\
&= \langle 0 \mid v\rangle\\
&= 0,
\end{aligned}$$

so $(A-\overline{\lambda}I)v = 0$ (by Exercise 9.1.9). This means $Av = \overline{\lambda}v$. Since Av is also equal to λv (and $v \neq 0$), this implies $\overline{\lambda} = \lambda$, so $\lambda \in \mathbb{R}$. □

Proof of Theorem 9.1.5. Let $m(x)$ be the minimal polynomial of A. From Lemma 9.1.11 and the Fundamental Theorem of Algebra (8.2.14), we know that $m(x)$ is a product of linear factors in $\mathbb{R}[x]$. Then all roots of $m(x)$ are real, so Exercise 9.1.7(3) implies that $m(x)$ has no repeated roots. We conclude from Proposition 8.3.7 that A is diagonalizable. □

(9.1.12) **Exercises.**

1) Show that the restriction to real matrices in Theorem 9.1.5 cannot be eliminated. More precisely, show that the symmetric matrix $\begin{bmatrix} 1 & i \\ i & -1 \end{bmatrix}$ is **not** diagonalizable (over $\mathbb{C}$).

   [*Hint:* What is the square of this matrix?]

2) Show that the eigenspaces of a real symmetric matrix are orthogonal to each other: if $Av = \lambda v$ and $Aw = \mu w$, with $\lambda \neq \mu$, then $v \cdot w = 0$.

   [*Hint:* Consider $v \cdot Aw$. You may use (without proof) the fact that the dot product is "bilinear" in the sense of Definition 9.2.1 below (see Example 9.2.2).]

(9.1.13) *Remark.* If A is a real symmetric $n \times n$ matrix, then, by combining Theorem 9.1.5 with Exercise 9.1.12(2) and Gram-Schmidt Orthogonalization (cf. Remark 9.4.14), one can show that $\mathbb{R}^n$ has an basis of eigenvectors that is orthonormal. This implies the converse of Exercise 9.1.3(7).

§9.2. Bilinear forms

In this section, we discuss a generalization of the dot product that applies to all vector spaces, not just $\mathbb{R}^n$.

(9.2.1) **Definition.** A ***bilinear form*** on the vector space V is a function $B\colon V \times V \to F$ that is linear in each variable, i.e.:

$$B(a_1v_1 + a_2v_2, w) = a_1\,B(v_1, w) + a_2\,B(v_2, w)$$

and

$$B(v, b_1w_1 + b_2w_2) = b_1\,B(v, w_1) + b_2\,B(v, w_2)$$

(9.2.2) **Example.** If we define $B(x, y) = x \cdot y$, then B is a bilinear form on $\mathbb{R}^n$. Note that if we identify x and y with column vectors $[x]$ and $[y]$ in the natural way, then

$$B(x, y) = x_1y_1 + x_2y_2 + \cdots + x_ny_n = [x_1, x_2, \ldots, x_n]\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = [x]^{\mathsf{T}}\,[y].$$

(9.2.3) **Exercise.** Let $\mathcal{B}$ be a basis of the n-dimensional vector space V.

1) If we define $B(x,y) = [x]_{\mathcal{B}}^{\mathsf{T}}\,[y]_{\mathcal{B}}$, then B is a bilinear form on V.
2) More generally, for any $M \in \mathrm{Mat}_{n\times n}(F)$, the formula $B_M(x,y) = [x]_{\mathcal{B}}^{\mathsf{T}} M\,[y]_{\mathcal{B}}$ defines a bilinear form on V. (The previous example is the special case where $M = I$.)
3) Conversely, every bilinear form on V is equal to B_M, for a unique $M \in \mathrm{Mat}_{n\times n}(F)$.
   [*Hint:* For B_M to agree with B on the basis vectors $v_1,\dots,v_n$, the matrix M must be $[B(v_i,v_j)]$. Then, since $\{v_i\}$ spans V, bilinearity implies $B_M = B$.]
4) Let B be a bilinear form on V. We say that B is ***nondegenerate*** if for every nonzero $v \in V$, there exists $w \in V$, such that $B(v,w) \neq 0$. Show that B_M is nondegenerate if and only if the matrix M is invertible.
5) Show that v and w can be interchanged in the definition of "nondegenerate." More precisely, for a bilinear form B on V, show that B is nondegenerate if and only if, for every nonzero $w \in V$, there exists $v \in V$, such that $B(v,w) \neq 0$.

(9.2.4) **Warning.** Exercise 9.2.3(5) is not true if V is allowed to be infinite-dimensional. (If $\dim V = \infty$, then the definition of ***nondegenerate*** should be revised to require both conditions.)

(9.2.5) **Exercises.** Assume U, V, W, Y, and Z are vector spaces.

1) Suppose $f\colon V \times W \to Y$ is bilinear and $T\colon Y \to Z$ is linear. Show that the composition $T \circ f\colon V \times W \to Z$ is bilinear.
2) Suppose that $f\colon V \times W \to U$ is bilinear and that $T_1\colon Y \to V$ and $T_2\colon Z \to W$ are linear. Show that the composition $f \circ (T_1 \times T_2)\colon (Y \times Z) \to U$ is bilinear, where
$$\big(f \circ (T_1 \times T_2)\big)(y,z) = f\big(T_1(y), T_2(z)\big) \quad \text{for } y \in Y \text{ and } z \in Z.$$
3) Suppose
   - f and g are bilinear functions from $V \times W$ to Z, and
   - $\{v_i\}$ and $\{w_j\}$ are bases of V and W, respectively,

   such that $f(v_i,w_j) = g(v_i,w_j)$ for all i and j. Show $f = g$.

§9.3. Dual space V^*

Suppose $\dim V = n$ and $\dim W = m$. Then, by choosing bases of V and W, it is easy to see that the vector space $\mathrm{Hom}(V,W)$ of all linear transformations from V to W is isomorphic to $\mathrm{Mat}_{m\times n}(F)$, so it has dimension mn. The case where $m = 1$ is particularly important:

(9.3.1) **Definition.**

1) A linear function $f\colon V \to F$ is called a ***linear functional*** on V.
2) Let $V^* = \{\text{ linear functionals on } V\}$. It is a vector space under the following operations (for $f,g \in V^*$ and $a \in F$):
   - $(f+g)(v) = f(v) + g(v)$ for $v \in V$.
   - $(af)(v) = a\,f(v)$ for $v \in V$.

   This is called the ***dual space*** of V.

(9.3.2) **Exercise.**

1) Suppose $v \in V$ and $f \in V^*$. Show that if $f(v) \neq 0$, then $V = Fv \oplus \ker f$.
2) Let $f,g \in V^*$. Show $\ker f = \ker g$ iff there is a nonzero scalar λ, such that $f = \lambda g$.
3) Show $(V_1 \oplus V_2)^*$ is naturally isomorphic to $V_1^* \oplus V_2^*$.
   [*Hint:* Define $\Phi\colon V_1^* \oplus V_2^* \to (V_1 \oplus V_2)^*$ by $\Phi(f_1,f_2)(v_1,v_2) = f_1(v_1) + f_2(v_2)$ and $\psi\colon (V_1 \oplus V_2)^* \to V_1^* \oplus V_2^*$ by $\psi(\mu) = (\psi_1(\mu),\psi_2(\mu))$, where $\psi_1(\mu)(v_1) = \mu(v_1,0)$ and $\psi_2(\mu)(v_2) = \mu(0,v_2)$. Show $\psi = \Phi^{-1}$. Your proof should not require the choice of a basis.]

If $\dim V < \infty$, then the above discussion shows that $\dim V^* = \dim V$, so $V^* \cong V$. Here is the usual proof of this fact:

(9.3.3) **Proposition.** *If V is finite-dimensional, then $V \cong V^*$.*

Proof. Fix a basis $\{v_1, \dots, v_n\}$ of V. Then, for each i, define $\hat{v}_i \in V^*$ by requiring that $\hat{v}_i(v_i) = 1$, but $\hat{v}_i(v_j) = 0$ for all $j \neq i$ (and extending to make it linear). It is not difficult to see that $\{\hat{v}_1, \dots, \hat{v}_n\}$ is a basis for V^*. (It is called the ***dual basis*** corresponding to $\{v_1, \dots, v_n\}$.) Namely, for any $f \in V^*$, $\sum_i f(v_i)\, \hat{v}_i$ is the unique linear combination that agrees with f on the basis $\{v_i\}$, and is therefore equal to f. This means that V^* has a basis with n elements, so $\dim V^* = n = \dim V$. Therefore $V \cong V^*$. □

Note that the above proof depends on choosing a basis of V. In fact, there is no canonical way to choose an isomorphism. However, instead of a basis, it suffices to choose only a bilinear form:

(9.3.4) **Exercise.** Suppose B is a nondegenerate bilinear form on V, and $\dim V < \infty$. Then there is an isomorphism $\hat{B}\colon V \to V^*$, defined by

$$(\hat{B}v)(w) = B(v, w) \quad \text{for } v, w \in V.$$

[*Hint:* Verify that $\hat{B}v \in V^*$, so $\hat{B}$ is a function from V to V^*. Then it is easy to check that $\hat{B}$ is linear. Use the nondegeneracy of B to show that $\ker \hat{B} = 0$. Since $\dim V = \dim V^* < \infty$, this implies that $\hat{B}$ is an isomorphism.]

(9.3.5) *Remark.* Conversely, if $\hat{B}\colon V \to V^*$ is any isomorphism, then a nondegenerate bilinear form B on V can be defined by the formula $B(v, w) = (\hat{B}v)(w)$. Hence, choosing a particular isomorphism from V to V^* is equivalent to choosing a nondegenerate bilinear form on V (when V is finite-dimensional).

(9.3.6) **Warning.** If V is infinite-dimensional, then $V \ncong V^*$, because $\dim V^* > \dim V$. (The reason is that if $\mathcal{B}$ is a basis of V, then V is isomorphic to the direct sum $\bigoplus_{v \in \mathcal{B}} F$ of infinitely many copies of F, but V^* is isomorphic to the direct product $\prod_{v \in \mathcal{B}} F$, which is much larger.)

(9.3.7) **Definition.** The dual space of the dual space is called the ***double-dual*** of V, and is denoted V^{**}.

If $\dim V < \infty$, then Proposition 9.3.3 implies $V^{**} = (V^*)^* \cong V^* \cong V$, so $V^{**} \cong V$. Although the isomorphisms from V^{**} to V^* and V to V^* both depend on the choice of a basis (or, at least, a bilinear form), the isomorphism from V to V^{**} does not require any such choice:

(9.3.8) **Proposition.** *If $\dim V < \infty$, then V^{**} is canonically isomorphic to V.*

Proof. For $v \in V$, define $\tilde{v} \in V^{**}$ by $\tilde{v}(f) = f(v)$, for $f \in V^*$. It is straightforward to verify that the map $v \mapsto \tilde{v}$ is linear. It is also quite easy to see that it is injective: for any basis of V that contains the nonzero vector v, we may let $\hat{v} \in V^*$ be the corresponding element of the dual basis, and then we have $\tilde{v}(\hat{v}) = \hat{v}(v) = 1$, so $\tilde{v} \neq 0$. Since $\dim V = \dim V^{**} < \infty$, this implies that the map is an isomorphism. □

(9.3.9) *Remark.* When V is infinite-dimensional, the above proof shows that the map $v \mapsto \tilde{v}$ is injective (because Zorn's Lemma provides a basis of V). However, it is not surjective (since $\dim V < \dim V^{**}$).

(9.3.10) **Definition.** For any linear operator $T\colon V \to V$, there is an associated linear transformation $T^{\mathsf{T}}\colon V^* \to V^*$ (sometimes called the ***transpose*** of T), defined by

$$(T^{\mathsf{T}} f)(v) = f(Tv) \quad \text{for all } v \in V \text{ and } f \in V^*.$$

When V is finite-dimensional, any nondegenerate bilinear form provides an identification of V^* with V, so the map $T^{\mathsf{T}}\colon V^* \to V^*$ can be thought of as a function from V to V:

(9.3.11) **Proposition.** *Suppose B is a nondegenerate bilinear form on the finite-dimensional vector space V, and $T\colon V \to V$ is linear. Then there is a unique linear transformation $T^*\colon V \to V$ (called the **adjoint** of T with respect to B), such that*

$$B(v, Tw) = B(T^* v, w) \quad \textit{for all } v, w \in V.$$

Proof. (existence) Let $T^{\mathsf{T}}\colon V^* \to V^*$ be the transpose of T. Exercise 9.3.4 provides an isomorphism $\widehat{B}\colon V \to V^*$, defined by $(\widehat{B}v)(w) = B(v,w)$, for $v, w \in V$. Let $T^* = \widehat{B}^{-1} \circ T^{\mathsf{T}} \circ \widehat{B}\colon V \to V$. Then, for any $v, w \in V$, we have

$$\begin{aligned} B(T^*v, w) &= (\widehat{B}T^*v)(w) && \text{(definition of } \widehat{B}) \\ &= (\widehat{B}(\widehat{B}^{-1} \circ T^{\mathsf{T}} \circ \widehat{B})v)(w) && \text{(definition of } T^*) \\ &= (T^{\mathsf{T}}(\widehat{B}v))(w) && \text{(cancel the inverse)} \\ &= (\widehat{B}v)(Tw) && \text{(definition of } T^{\mathsf{T}}) \\ &= B(v, Tw) && \text{(definition of } \widehat{B}). \end{aligned}$$ □

(uniqueness) Suppose T' is another adjoint of T. Then, for all $v, w \in V$, we must have $B(T'v, w) = B(T^*v, w)$, so the bilinearity of B implies $B(T'v - T^*v, w) = 0$. Since B is nondegenerate, we conclude that $T'v - T^*v = 0$, so $T'v = T^*v$.

Alternate proof of existence. Assume, for simplicity, that $V = F^n$, so there is an invertible matrix $M \in \mathrm{Mat}_{n\times n}(F)$, such that $B(v,w) = v^{\mathsf{T}}Mw$ for all $v, w \in F^n$ (cf. Exercise 9.2.3(3)). Given $T \in \mathrm{Mat}_{n\times n}(F)$, let $T^* = (MTM^{-1})^{\mathsf{T}}$, so

$$B(v, Tw) = v^{\mathsf{T}}MTw = v^{\mathsf{T}}MTM^{-1}Mw = v^{\mathsf{T}}(T^*)^{\mathsf{T}}Mw = (T^*v)^{\mathsf{T}}Mw = B(T^*v, w),$$

as desired. □

(9.3.12) **Exercise.** For the (standard) dot product on $\mathbb{R}^n$, show that the adjoint of any matrix A is its ordinary matrix transpose.

[*Hint:* $B(v,w) = v^{\mathsf{T}}w$. Use the fact that the transpose of a product is the product of the transposes *in the reverse order*.]

(9.3.13) **Exercise.** Let

- B be a nondegenerate bilinear form on V, with $\dim V < \infty$, and
- S and T be linear operators on V.

Show:

1) $(S+T)^* = S^* + T^*$.
2) $(aT)^* = aT^*$, for all $a \in F$.
3) $(ST)^* = T^*S^*$.
4) If B is ***symmetric*** (which means that $B(v,w) = B(w,v)$ for all $v, w \in V$), then $T^{**} = T$ (where T^{**} means $(T^*)^*$).
5) If T is invertible, then T^* is also invertible, and $(T^*)^{-1} = (T^{-1})^*$.

(9.3.14) *Remark.* The adjoint T^* may not exist when V is infinite-dimensional. For example, suppose V has a basis $\{v_i\}_{i=1}^{\infty}$, such that $B(v_i, v_j) = 0$ for $i \neq j$, but $B(v_1, v_1) \neq 0$. Define $T\colon V \to V$ by $Tv_i = v_i + v_1$, and suppose $T^*v_1 = \sum_{i=1}^{n} a_i v_i$. Then

$$B(Tv_{n+1}, v_1) = B(v_{n+1} + v_1, v_1) = B(v_{n+1}, v_1) + B(v_1, v_1) = 0 + B(v_1, v_1) \neq 0,$$

but

$$B(v_{n+1}, T^*v_1) = B(v_{n+1}, \textstyle\sum_{i=1}^{n} a_i v_i) = \sum_{i=1}^{n} a_i\, B(v_{n+1}, v_i) = \sum_{i=1}^{n} a_i\, 0 = 0.$$

This contradicts the fact that $B(Tv_{n+1}, v_1) = B(v_{n+1}, T^*v_1)$, so the adjoint T^* does not exist.

§9.4. Hermitian forms and diagonalizability

The ordinary dot product on $\mathbb{R}^n$ has the nice property that if $v \neq 0$, then $v \cdot v \neq 0$. In fact, $v \cdot v$ is a positive number (unless $v = 0$), so it has a square root, which is the ***length*** (or "***norm***") of v:

$$\|v\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} = \sqrt{v \cdot v}.$$

Unfortunately, there is no such bilinear form on $\mathbb{C}^n$ (unless $n = 1$):

(9.4.1) **Exercise.** Suppose B is any bilinear form on $\mathbb{C}^n$, and $n \geq 2$. Show there exists some nonzero $v \in \mathbb{C}^n$, such that $B(v, v) = 0$.

[*Hint:* If $v, w \in \mathbb{C}^n$ with $B(w, w) \neq 0$, then $B(v + tw, v + tw)$ is a quadratic polynomial in t.]

We can eliminate this problem by replacing bilinear forms with Hermitian forms:

(9.4.2) **Definition.** Suppose V is a vector space over $\mathbb{C}$.

1) A function $B\colon V \times V \to \mathbb{C}$ is a ***Hermitian form*** if it is linear in the first argument and conjugate-symmetric. For $u, v, w \in V$ and $s, t \in \mathbb{C}$, this means:
   - $B(su + tv, w) = sB(u, w) + tB(v, w)$, and
   - $B(v, w) = \overline{B(w, v)}$, where $\overline{z}$ denotes the complex conjugate $a - bi$ of the complex number $z = a + bi$.
2) A Hermitian form B on V is ***nondegenerate*** if for every nonzero $v \in V$, there exists $w \in V$, such that $B(v, w) \neq 0$.

(9.4.3) **Exercise.** Show that any Hermitian form is conjugate linear in the second argument. (This means $B(w, su + tv) = \overline{s}B(w, u) + \overline{t}B(w, v)$.)

(9.4.4) **Warning.** Many authors (including physicists) reverse the variables: they require a Hermitian form to be linear in the second argument (and conjugate-linear in the first).

(9.4.5) **Examples.**

1) We have the standard Hermitian form on $\mathbb{C}^n$ (cf. Definition 9.1.8):
$$B(v, w) = v_1\,\overline{w_1} + v_2\,\overline{w_2} + \cdots + v_n\,\overline{w_n}$$
for $v = (v_1, \ldots, v_n)$ and $w = (w_1, \ldots, w_n)$. Note that $B(v, v) > 0$ for all nonzero v.
2) A different Hermitian form B' on $\mathbb{C}^n$ can be defined by
$$B'(v, w) = v_1\,\overline{w_1} + v_2\,\overline{w_2} + \cdots + v_{n-1}\,\overline{w_{n-1}} - v_n\,\overline{w_n}.$$
Note that $B'(v, v)$ can be negative (because of the minus sign in front of $v_n\,\overline{w_n}$).

(9.4.6) *Remark.* The ***length*** (or "***norm***") of vector v in $\mathbb{C}^n$ is often defined to be
$$\|v\| = \sqrt{|v_1|^2 + |v_2|^2 + \cdots + |v_n|^2} = \sqrt{v_1\,\overline{v_1} + v_2\,\overline{v_2} + \cdots + v_n\,\overline{v_n}} = \sqrt{B_1(v, v)},$$
for the standard Hermitian form B of Example 9.4.5(1).

We have the following analogue of Proposition 9.3.11:

(9.4.7) **Exercises.** Suppose B is a nondegenerate Hermitian form on V, with $\dim V < \infty$, and $T\colon V \to V$ is linear.

1) Show there is a unique linear transformation $T^*\colon V \to V$ (called the ***adjoint*** of T with respect to B), such that
$$B(v, Tw) = B(T^*v, w) \quad \text{for all } v, w \in V.$$
2) Show $T^{**} = T$.
   [*Hint:* Use the fact that B is conjugate-symmetric.]
3) For the standard Hermitian form B on $\mathbb{C}^n$ that is defined in Example 9.4.5(1), show that the adjoint of any matrix A is its conjugate-transpose.
   [*Hint:* $B(v, w) = v^{\mathsf{T}}\,\overline{w}$.]

(9.4.8) **Warning.** If $\dim V < \infty$, then any nondegenerate bilinear form provides an isomorphism from V to V^*. However, a nondegenerate Hermitian form will provide an isomorphism from V not to V^*, but to the conjugate $\overline{V^*}$ of V^*. (The ***conjugate*** of a vector space V has the same vector addition, but uses a different scalar multiplication $*$ that is defined by $\alpha * v = \overline{\alpha}v$.) However, this is not too serious a problem, because every vector space is isomorphic to its conjugate. (It is easy to see that every basis of V is also a basis of $\overline{V}$, so $\dim V = \dim \overline{V}$.)

(9.4.9) **Definition.** Fix a nondegenerate Hermitian form B on V, and suppose $T\colon V \to V$ is a linear operator.

- We say that T is ***self-adjoint*** if $T^* = T$.
- We say that T is ***unitary*** if $T^* = T^{-1}$.
- We say that T is ***normal*** if T commutes with its adjoint (i.e., $T^*T = TT^*$).

Note that if T is either self-adjoint or unitary, then it is normal.

(9.4.10) **Exercise.** Show that T is unitary if and only if $B(Tv, Tw) = B(v, w)$ for all $v, w \in V$.

(9.4.11) **Definition.** Suppose B is a Hermitian form on V.

1) The ***orthogonal complement*** of a subspace W of V is
$$W^\perp = \{\, v \in V \mid B(v, w) = 0 \text{ for all } w \in W \,\}.$$
This is a subspace of V. (*Why?*)
2) B is said to be ***positive-definite*** if $B(v, v) > 0$ for all nonzero $v \in V$.

(9.4.12) **Exercise.** Assume that B is a positive-definite Hermitian form on V, that W is a subspace of V, and that $T\colon V \to V$ is a linear operator.

1) Show B is nondegenerate.
2) Show $V = W \oplus W^\perp$ if V is finite-dimensional.
3) Assume W is T-invariant.
   (a) Show that if T is self-adjoint, then $W^\perp$ is T-invariant.
   [*Hint:* It is easy to see that $W^\perp$ is T^*-invariant.]
   (b) Show that if T is unitary and $\dim V < \infty$, then $W^\perp$ is T-invariant.
   More generally, if T is normal (and $\dim V < \infty$), then $W^\perp$ is T-invariant, but we will not prove this until Exercise 9.4.19 below, since it is more difficult.

(9.4.13) **Definition.** A basis $\{v_1, \ldots, v_n\}$ is ***orthonormal*** (with respect to the Hermitian form B) if $B(v_i, v_i) = 1$ for all i, but $B(v_i, v_j) = 0$ whenever $i \neq j$.

(9.4.14) *Remark.* The Gram-Schmidt Orthogonalization process is an inductive procedure for modifying any set of vectors in $\mathbb{R}^n$ to obtain an orthonormal basis of the span of those vectors. It implies that every real vector space has an orthonormal basis. The same idea can be used to show that if B is any Hermitian form on any complex vector space V, then V has an orthonormal basis (with respect to B).

(9.4.15) **Theorem** (Spectral Theorem for self-adjoint operators)**.** *If $\dim V < \infty$, then every self-adjoint linear operator on V is diagonalizable.*

More precisely, assume:

- *B is a positive-definite Hermitian form on a finite-dimensional vector space V over $\mathbb{C}$, and*
- *$T\colon V \to V$ is a self-adjoint linear operator.*

Then T is diagonalizable.

In fact, V has an orthonormal basis consisting of eigenvectors of V.

Proof. Since $\mathbb{C}$ is algebraically closed, we know that the characteristic polynomial of T has a root, so T has an eigenvector w. (Multiplying by a scalar, we may assume $B(w, w) = 1$.) Let $W = \mathbb{C}w$. Then W is clearly T-invariant, so Exercise 9.4.12(3a) tells us that $W^\perp$ is also T-invariant.

Let $T^\perp$ and $B^\perp$ be the restrictions of T and B to $W^\perp$. Then $B^\perp$ is a positive-definite Hermitian form and $T^\perp$ is a self-adjoint linear transformation. (Why?) Therefore, by induction on $\dim V$, we may assume there is an (orthonormal) basis $v_1, \ldots, v_{n-1}$ of $W^\perp$ that consists of eigenvectors of $T^\perp$. Then $\{v_1, \ldots, v_{n-1}, w\}$ is an (orthonormal) basis of V that consists of eigenvectors of T. □

(9.4.16) **Exercise.** Let $U \in \mathrm{Mat}_{n\times n}(\mathbb{C})$, and let B be the standard Hermitian form on $\mathbb{C}^n$ (defined in Example 9.4.5(1)). Show that the following are equivalent:

1) The map $v \mapsto Uv$ is a unitary operator (with respect to the Hermitian form B).
2) U^{-1} is the conjugate-transpose of U.

3) For every orthonormal basis $\{v_1, \dots, v_n\}$ of $\mathbb{C}^n$ (with respect to the Hermitian form B), the set $\{Uv_1, \dots, Uv_n\}$ is also an orthonormal basis.

If these conditions hold, the matrix U is said to be ***unitary***.

(9.4.17) **Corollary.** *Suppose A is a self-adjoint matrix in* $\mathrm{Mat}_{n\times n}(\mathbb{C})$ (*which means that A is equal to its conjugate-transpose*). *Then there is a unitary matrix U, such that UAU^{-1} is diagonal.*

All normal operators, not only the self-adjoint ones, are diagonalizable by an orthonormal basis. However, the proof takes a little more work, because it is not so obvious in this case that the orthogonal complement of an invariant subspace is invariant. The following exercise outlines the proof.

(9.4.18) **Exercise.** Assume:

- B is a positive-definite Hermitian form on a finite-dimensional vector space V,
- $T\colon V \to V$ is a normal linear operator, and
- $\lambda, \mu \in \mathbb{C}$.

Show:

1) $\ker T = \ker T^*$.
   [*Hint:* $v \in \ker T \Rightarrow 0 = B(Tv, Tv) = B(T^*Tv, v) = B(TT^*v, v) = B(T^*v, T^*v)$.]
2) If $Tv = \lambda v$, then $T^*v = \overline{\lambda}v$.
   [*Hint:* Apply (1) to $T - \lambda I$.]
3) If $Tv = \lambda v$ and $Tw = \mu w$, with $\lambda \neq \mu$, then $B(v, w) = 0$.
   [*Hint:* $\lambda B(v, w) = B(Tv, w) = B(v, T^*w) = B(v, \overline{\mu}w) = \mu B(v, w)$.]
4) If $T^k v = 0$, for some $k \in \mathbb{N}$, then $Tv = 0$.
   [*Hint:* $0 = B(T^2v, T^2v) = B(T^*Tv, T^*Tv) \Rightarrow T^*Tv = 0 \Rightarrow 0 = B(T^*Tv, v) = B(Tv, Tv)$.]
5) T is diagonalizable.
   [*Hint:* Applying (4) to $T - \lambda I$ shows that the minimal polynomial of T has no repeated roots.]
6) V has an orthonormal basis consisting of eigenvectors of T.
   [*Hint:* Choose an orthonormal basis of each eigenspace of T, and let $\mathcal{B}$ be the union of all these vectors. Then (5) implies that $\mathcal{B}$ spans V, and (3) implies that all the vectors in $\mathcal{B}$ are orthogonal to each other.]

(9.4.19) **Exercise.** Assume

- B is a positive-definite Hermitian form on a finite-dimensional vector space V,
- $T\colon V \to V$ is a normal linear operator, and
- W is a T-invariant subspace of V.

Show that $W^\perp$ is T-invariant.

[*Hint:* You may assume Exercise 9.4.18.]

Exercise 9.4.18 implies Theorem 9.1.5:

(9.4.20) **Exercise.** Let $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$, such that $A^{\mathsf{T}} = A$. (I.e., A is ***symmetric***.) Show that A is diagonalizable (over $\mathbb{R}$).

[*Hint:* Exercise 9.4.18(2) shows that the eigenvalues of A are real.]

§9.5. Tensor products of vector spaces (advanced)

(9.5.1) **Example.** Consider $F[x]$ and $F[y]$, the polynomial rings over F with two different variables x and y. Each of these is a vector space over F, and we can multiply elements of $F[x]$ by elements of $F[y]$. For example, taking $2 + 5x \in F[x]$ and $3 - y + 4y^2 \in F[y]$, we have

$$(2 + 5x)(3 - y + 4y^2) = 6 + 15x - 2y - 5xy + 8y^2 + 20xy^2.$$

Note that the result of such a multiplication is (usually) <u>not</u> in $F[x]$ or $F[y]$, but in the vector space $F[x, y]$ of polynomials in two variables.

The tensor product is an idea that generalizes this, allowing the elements of any vector space V to be multiplied by the elements of any other vector space W. For $v \in V$ and $w \in W$, the result of multiplying v times w is written $v \otimes w$, and belongs to a vector space $V \otimes W$, called the *tensor product* of V and W.

For example, $F[x] \otimes F[y] = F[x, y]$. Note that

- $\{x^i\}$ is a basis of $F[x]$,
- $\{y^j\}$ is a basis of $F[y]$, and
- $\{x^i y^j\}$ is a basis of $F[x, y]$.

Therefore, if $V = F[x]$ and $W = F[y]$, then a basis of $V \otimes W$ can be obtained by multiplying every element of a basis of V by every element of a basis of W. This is true in general:

(9.5.2) **Definition.** Suppose $\{v_i\}$ and $\{w_j\}$ are bases of V and W, respectively. Then the ***tensor product*** $V \otimes W$ is the vector space whose basis is the set of all symbols $v_i \otimes w_j$. Furthermore, for

$$v = \sum \alpha_i v_i \in V \text{ and } w = \sum \beta_j w_j \in W,$$

we define

$$v \otimes w = \sum \alpha_i \beta_j (v_i \otimes w_j).$$

(9.5.3) **Observation.** *Since $\{v_i \otimes w_j\}$ is a basis of $V \otimes W$, it is obvious that*

$$\dim(V \otimes W) = (\dim V) \cdot (\dim W).$$

(9.5.4) **Warning.** Not every element of $V \otimes W$ is of the form $v \otimes w$ (such elements are called "***pure***" tensors or "decomposable" tensors), unless either $\dim V$ or $\dim W$ is 0 or 1 (see Exercise 9.5.14(2)). Instead, most elements are a sum of elements of this form. For example, most polynomials in $F[x, y]$ (such as $1 + xy$) are not of the form $f(x)\, g(y)$.

The following straightforward exercise shows that this multiplication of vectors is distributive (and that scalars can be pulled out):

(9.5.5) **Exercise.** Show that the map $(v, w) \mapsto v \otimes w$ is bilinear. More precisely, for $v, v_1, v_2 \in V$, $w, w_1, w_2 \in W$, and $\alpha \in F$, we have:

a) $(v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w$,

b) $v \otimes (w_1 + w_2) = v \otimes w_1 + v \otimes w_2$, and

c) $(\alpha v) \otimes w = \alpha(v \otimes w) = v \otimes (\alpha w)$.

(9.5.6) **Exercise.** Show that if v and w are nonzero vectors in V and W, respectively, then the element $v \otimes w$ of $V \otimes W$ is nonzero.

(9.5.7) *Remarks.*

1) It can be shown that $V \otimes W$ does not depend on the choice of the bases $\{v_i\}$ and $\{w_i\}$ (cf. Corollary 9.5.11). In fact, there are alternative definitions of $V \otimes W$ that do not use a basis at all. In particular, some textbooks take the "universal" property in Proposition 9.5.9 below as the definition of $V \otimes W$. Others define $V \otimes W$ to be the vector space spanned by the symbols $v \otimes w$, for $v \in V$ and $w \in W$, subject only to relations (in 9.5.5(a,b,c)) that make the map $(v, w) \mapsto v \otimes w$ bilinear.

2) Although we have defined the tensor product only for vector spaces, it can actually be defined for modules over any (commutative) ring. Since most modules do not have a basis, Definition 9.5.2 cannot be used, so one of the other approaches mentioned in (1) is needed for the generalization.

   Warning. In the setting of rings and modules, Exercise 9.5.6 is not usually valid: there often exist nonzero $m \in M$ and $n \in N$, such that $m \otimes n = 0$ in $M \otimes N$.

(9.5.8) **Exercise.** Show that if $\{v_i'\}$ spans V and $\{w_j'\}$ spans W, then $\{v_i' \otimes w_j'\}$ spans $V \otimes W$.

[*Hint:* Given $v \in V$ and $w \in W$, write $v = \sum \alpha_i v_i'$ and $w = \sum \beta_j w_j'$. Then use bilinearity to show that $v \otimes w$ is in the span of $\{v_i' \otimes w_j'\}$.]

The following fundamental property is usually much more useful than Definition 9.5.2. More precisely, if you remember this proposition (and other basic facts, such as Exercises 9.5.5 and 9.5.8), then you will almost never need (or want) Definition 9.5.2.

(9.5.9) **Proposition** (Universal property of the tensor product)**.** *Assume V, W, and Z are vector spaces over F. For any bilinear function $f\colon V \times W \to Z$, there is a unique linear map $T\colon V \otimes W \to Z$, such that $T(v \otimes w) = f(v, w)$ for all $v \in V$ and $w \in W$.*

Proof. Since $\{v_i \otimes w_j\}$ is a basis, there is a unique linear transformation $T\colon V \otimes W \to Z$, such that $T(v_i \otimes w_j) = f(v_i, w_j)$ for all i and j. (This establishes the uniqueness.) Now, given $v = \sum \alpha_i v_i \in V$ and $w = \sum \beta_j w_j \in W$, combining the bilinearity of $\otimes$ and f with the linearity of T yields

$$T(v \otimes w) = \sum \alpha_i \beta_j \, T(v_i \otimes w_j) = \sum \alpha_i \beta_j \, f(v_i \otimes w_j) = f(v \otimes w),$$

as desired. □

(9.5.10) *Remark.* Similar universal properties are common in algebra: saying an object U is "universal" means that every object O with a certain property is a quotient of that object (or, equivalently, that there is a homomorphism from U to O). For example, we have seen that every group with an n-element generating set is a quotient of the free group on n generators, so free groups are universal. Similarly, free modules are universal, because every module is a quotient of a free module.

Two examples we have already seen are:

1) Suppose G is the free group on a set X. If f is any function from X to a group H, then there is a unique homomorphism $\varphi\colon G \to H$, such that $\varphi(x) = f(x)$ for all $x \in X$ (see Theorem 4.2.9)
2) Suppose S is a basis for a free module M over R. If f is any function from S to an R-module N, then there is a unique homomorphism $\varphi\colon M \to N$, such that $\varphi(s) = f(s)$ for all $s \in S$ (see Exercise 6.4.7(4)).

The following fundamental result is a strengthening of Exercise 9.5.6. It shows that (up to isomorphism) $V \otimes W$ is independent of the choice of basis.

(9.5.11) **Corollary.** *If $\{v_i'\}$ and $\{w_j'\}$ are any bases of V and W, respectively, then $\{v_i' \otimes w_j'\}$ is a basis of $V \otimes W$.*

Proof. From Exercise 9.5.8, it suffices to show that $\{v_i \otimes w_j\}$ is linearly independent. Let $\{\widehat{v}_i\}$ and $\{\widehat{w}_j\}$ be the linear functionals in V^* and W^* that are dual to the bases $\{v_i\}$ and $\{w_j\}$ (as in the proof of Proposition 9.3.3), and fix some i_0 and j_0. Then there is a bilinear function $f\colon V \times W \to F$ defined by $f(v, w) = \widehat{v}_{i_0}(v)\, \widehat{w}_{j_0}(w)$. Let $T\colon V \otimes W \to F$ be the corresponding linear map provided by the universality of the tensor product. If $\sum \alpha_{i,j}\, v_i \otimes w_j = 0$, then

$$0 = T(0) = T\left(\sum \alpha_{i,j} v_i \otimes w_j\right) = \sum \alpha_{i,j}\, \widehat{v}_{i_0}(v_i)\, \widehat{w}_{j_0}(w_j) = \alpha_{i_0,j_0} \cdot 1 \cdot 1 = \alpha_{i_0,j_0}.$$

Since i_0 and j_0 are arbitrary, we conclude that $\alpha_{i,j} = 0$ for all i and j. So $\{v_i \otimes w_j\}$ is linearly independent. □

(9.5.12) **Exercise.** Show that if $\{v_i'\}$ and $\{w_j'\}$ are linearly independent subsets of V and W, respectively, then $\{v_i' \otimes w_j'\}$ is a linearly independent subset of $V \otimes W$.

[*Hint:* Extend $\{v_i'\}$ and $\{w_j'\}$ to bases and apply Corollary 9.5.11.]

(9.5.13) **Example.** We can use the universal property to show that $F[x] \otimes F[y] \cong F[x, y]$.

Proof. The function $B\colon F[x] \times F[y] \to F[x, y]$, defined by $B(f(x), g(y)) = f(x)\, g(y)$, is obviously bilinear, so the universality of the tensor product implies there is a corresponding linear map $T\colon F[x] \otimes F[y] \to F[x, y]$. Since it maps the basis $\{x^i \otimes y^j\}$ of $F[x] \otimes F[y]$ to the basis $\{x^i y^j\}$ of $F[x, y]$, T is an isomorphism. □

(9.5.14) **Exercises.**

1) Suppose $\{w_i\}$ is a basis of W. Show that every element of $V \otimes W$ can be written uniquely in the form $\sum v_i \otimes w_i$, where $v_i \in V$.
2) Assume $\{v_1, v_2\}$ and $\{w_1, w_2\}$ are pairs of linearly independent vectors in V and W, respectively. Show that $v_1 \otimes w_1 + v_2 \otimes w_2$ is not a pure tensor.
   [*Hint:* For $v = \sum_i a_i v_i$ and $w = \sum_j b_j w_j$, use Corollary 9.5.11 to show that $v \otimes w \neq v_1 \otimes w_1 + v_2 \otimes w_2$.]
3) Is the image of a bilinear function $f: V \times W \to Z$ always a subspace of Z?
4) Suppose $v_1, v_2 \in V$ and $w_1, w_2 \in W$, such that $v_1 \otimes w_1 \neq 0$. Show that $v_1 \otimes w_1 = v_2 \otimes w_2$ iff there exists some nonzero $\alpha \in F$, such that $v_2 = \alpha v_1$ and $w_2 = (1/\alpha) w_1$.
5) Suppose V' and W' are subspaces of V and W, respectively.
   (a) Show that $V' \otimes W'$ can naturally be identified with a subspace of $V \otimes W$. More precisely, show that if $T: V' \otimes W' \to V \otimes W$ is the unique linear transformation such that $T(v \otimes w) = v \otimes w$ for $v \in V'$ and $w \in W'$, then T is an isomorphism onto its image.
   (b) Show that if $V' \otimes W' = V \otimes W$, then $V' = V$ and $W' = W$.
6) Show that the tensor product of vector spaces is commutative, distributive, and associative:
   (a) $V \otimes W$ is naturally isomorphic to $W \otimes V$,
   (b) $U \otimes (V \oplus W)$ is naturally isomorphic to $(U \otimes V) \oplus (U \otimes W)$, and
   (c) $(U \otimes V) \otimes W$ is naturally isomorphic to $U \otimes (V \otimes W)$.

By comparing dimensions, we see that $F^m \otimes F^n$ is isomorphic to F^{mn}, and it is also isomorphic to $\mathrm{Mat}_{mn}(F)$. Since $\mathrm{Mat}_{mn}(F)$ can be identified with the vector space $\mathrm{Hom}_F(F^m, F^n)$ of all linear transformations from F^m to F^n, and $F^m \cong (F^m)^*$, this implies that $(F^m)^* \otimes F^n \cong \mathrm{Hom}_F(F^m, F^n)$. We have the following generalization:

(9.5.15) **Exercise.** Show that if $\dim W < \infty$, then $V^* \otimes W$ is naturally isomorphic to $\mathrm{Hom}_F(V, W)$.
[*Hint:* There is a linear transformation $T: V^* \otimes W \to \mathrm{Hom}_F(V, W)$, such that $T(f \otimes w)(v) = f(v)\, w$. Choose a basis $\{w_1, \ldots, w_n\}$ of W, and let $\{\widehat{w}_1, \ldots, \widehat{w}_n\}$ be the dual basis. To show T is onto, note that $\widehat{w}_i \circ S \in V^*$, for any $S \in \mathrm{Hom}_F(V, W)$. To show T is one-to-one, note that any element of $V^* \otimes W$ can be written in the form $\sum f_i \otimes w_i$.]

(9.5.16) *Remark.* More generally, without any assumption on $\dim W$, it is not difficult to see that $V^* \otimes W$ is naturally isomorphic to the subspace $\mathrm{Hom}_F^{\text{f.d.}}(V, W)$ of $\mathrm{Hom}_F(V, W)$ consisting of the linear transformations with finite-dimensional range.

(9.5.17) **Exercise.** There is a natural linear map $\tau: V^* \otimes V \to F$, defined by $\tau(f \otimes v) = f(v)$. Show that if V is finite-dimensional, and $V^* \otimes V$ is identified with $\mathrm{Hom}_F(V, V)$ in the natural way (see Exercise 9.5.15), then $\tau(T)$ is the trace of T, for all $T \in \mathrm{Hom}_F(V, V)$.

(Recall that if $A = (a_{i,j}) \in \mathrm{Mat}_{n \times n}$, then the ***trace*** of A is the sum of the diagonal entries of A: $\mathrm{trace}\, A = a_{1,1} + \cdots + a_{n,n}$. And the trace of a linear operator $T: V \to V$ is defined by choosing a basis of V, and letting T be the trace of the matrix corresponding to T. This exercise provides a different definition of the trace of a linear transformation, without any need to choose a basis.)

The following fundamental result shows that the tensor product is completely determined (up to isomorphism) by the universal property in Proposition 9.5.9. Therefore, this universality can be (and often is!) taken as the definition of the tensor product.

(9.5.18) **Corollary.** *Assume U, V, W, and Z are vector spaces over F, and that $\boxtimes: V \times W \to U$ is bilinear. Also assume that, for any bilinear function $f: V \times W \to U$, there is a unique linear map $T: U \to Z$, such that $T(v \boxtimes w) = f(v, w)$ for all $v \in V$ and $w \in W$. Then there is a unique isomorphism $\Phi: V \otimes W \xrightarrow{\cong} U$, such that $\Phi(v \otimes w) = v \boxtimes w$, for all $(v, w) \in V \times W$.*

Proof. Since $\boxtimes$ is bilinear, the universality of the tensor product implies there is a linear map $\Phi: V \otimes W \to U$, such that $\Phi(v \otimes w) = v \boxtimes w$, for all $(v, w) \in V \times W$. Similarly, since $(v, w) \mapsto v \otimes w$ is bilinear, the assumed universality of U implies there is a linear map $\psi: U \to V \otimes W$, such that $\psi(v \boxtimes w) = v \otimes w$, for all $(v, w) \in V \times W$. Then, for all $(v, w) \in V \times W$, we have

$$(\psi \circ \Phi)(v \otimes w) = \psi\big(\Phi(v \otimes w)\big) = \psi(v \boxtimes w) = v \otimes w.$$

However, the uniqueness in Proposition 9.5.9 implies that the identity map is the unique linear map $T\colon V \otimes W \to V \otimes W$, such that $T(v \otimes w) = v \otimes w$. So $\psi \circ \Phi$ must be the identity map on $V \otimes W$. Similarly, the uniqueness of maps $U \to Z$ (with $Z = U$) implies that $\Phi \circ \psi$ must be the identity map on U. Hence, Φ is an isomorphism. □

(9.5.19) **Exercises.**

1) There is a natural homomorphism $\Lambda\colon V^* \otimes W^* \to (V \otimes W)^*$ that satisfies
$$\big(\Lambda(f \otimes g)\big)(v \otimes w) = f(v)\, g(w).$$
   (a) Show that Λ is one-to-one.
   (b) Show that Λ is onto if and only if either V or W is finite-dimensional.
2) Generalizing (1), show that if V, W, and Z are vector spaces, then $\mathrm{Hom}_F(V \otimes W, Z)$ is naturally isomorphic to $\mathrm{Hom}_F\big(V, \mathrm{Hom}_F(W, Z)\big)$.
   [*Hint:* Given $f \in \mathrm{Hom}(V \otimes W, Z)$, any fixed $v \in V$ yields a homomorphism from W to Z. Conversely, any element of $\mathrm{Hom}\big(V, \mathrm{Hom}_F(W, Z)\big)$ yields a bilinear function from $V \times W$ to Z.]
3) Suppose $T_1\colon V_1 \to W_1$ and $T_2\colon V_2 \to W_2$ are linear transformations.
   (a) Show there is a unique linear transformation $T_1 \otimes T_2\colon V_1 \otimes V_2 \to W_1 \otimes W_2$, such that $(T_1 \otimes T_2)(v_1 \otimes v_2) = T_1(v_1) \otimes T_2(v_2)$ for all $v_1 \in V_1$ and $v_2 \in V_2$.
   (b) Show the image of $T_1 \otimes T_2$ is $T_1(V_1) \otimes T_2(V_2)$.
   (c) Show $\ker(T_1 \otimes T_2) = (\ker T_1) \otimes V_2 + V_1 \otimes (\ker T_2)$.
   (d) Assume $V_1 = W_1$ and $V_2 = W_2$, so T_1 and T_2 are linear operators. Show that if T_1 and T_2 are diagonalizable, then $T_1 \otimes T_2$ is diagonalizable.
   (e) The map $(T_1, T_2) \mapsto T_1 \otimes T_2$ is bilinear, so it yields a linear transformation
$$\Lambda\colon \mathrm{Hom}_F(V_1, W_1) \otimes \mathrm{Hom}_F(V_2, W_2) \to \mathrm{Hom}_F(V_1 \otimes V_2, W_1 \otimes W_2).$$
   Show that if V_1, V_2, W_1, and W_2 are finite-dimensional, then Λ is an isomorphism.

Chapter 10

Jordan Canonical Form

(10.0.1) **Notation.**

- F is a field,
- $F[x]$ is the ring of polynomials over F, and
- $\mathrm{Mat}_{n\times n}(F)$ is the ring of $n \times n$ matrices with entries in F.

For $A_1, A_2 \in \mathrm{Mat}_{n\times n}(F)$, recall that A_1 is ***similar*** to A_2 if there exists an invertible matrix $P \in \mathrm{Mat}_{n\times n}(F)$, such that $PA_1P^{-1} = A_2$. (This is an equivalence relation.) Another way of saying this is that A_1 and A_2 represent the same linear transformation from F^n to F^n, but with respect to different bases.

We would like a nice way to find all of the different possible $n \times n$ matrices, up to similarity. This will provide us with an easy way to decide whether two matrices are similar, and will also tell us all of the different linear transformations from F^n to itself, up to a change of basis.

Several solutions to this problem are known, but we will only talk about two of the most famous ones: Jordan Canonical Form and Rational Canonical Form.

§10.1. General method for canonical forms

For each $A \in \mathrm{Mat}_{n\times n}(F)$, let M_A be the corresponding $F[x]$-module. Namely, $M_A = F^n$ with the usual vector addition, but with scalar multiplication defined by

$$f(x) \cdot v = f(A)\, v = \sum_{i=0}^{m} a_i A^i v \qquad \text{for } f(x) = \sum_{i=0}^{m} a_i x^i \in F[x].$$

(10.1.1) **Exercise.** M_A is a finitely generated, torsion module.
[*Hint:* The vector space F^n is finite-dimensional.]

(10.1.2) **Exercise.** $M_{A_1} \cong M_{A_2}$ iff A_1 is similar to A_2.

The above observation translates our question about similarity of matrices to a problem about isomorphism of modules, which we can solve by using the structure of modules over a PID. (Recall that $F[x]$ is a PID, since it is Euclidean.)

(10.1.3) **Exercise.** Show $M_{A\oplus B} \cong M_A \oplus M_B$ (see Definition 8.2.5).

By the structure of modules over a PID, we may write M_A as a direct sum

$$M_A \cong M_1 \oplus \cdots \oplus M_r, \text{ where each } M_i \text{ is a nonzero, cyclic, primary module}$$

(and the M_i's are unique, up to a reordering). To construct a canonical form, we choose a matrix A_i, such that $M_i \cong M_{A_i}$, for then A is similar to $A_1 \oplus \cdots \oplus A_r$. However, in order to be *canonical*, the choice of A_i must be done systematically, so that A is similar to some other given matrix A' iff $A_i = A'_i$ for all i (up to reordering the factors).

More precisely, we choose A_i to be in a predetermined set $\mathcal{A}$ of matrices, such that each cyclic, primary module is isomorphic to M_A for some *unique* $A \in \mathcal{A}$.

(10.1.4) **Exercise.** Let $\mathcal{A} \subseteq \bigcup_{n=1}^{\infty} \mathrm{Mat}_{n\times n}(F)$, such that

a) M_A is a cyclic, primary module, for each $A \in \mathcal{A}$, and, conversely,

b) each cyclic, primary $F[x]$-module C is isomorphic to the module M_A for some *unique* $A = A_C \in \mathcal{A}$.

For each $A \in \mathrm{Mat}_{n\times n}(F)$, show:

1) A is similar to $A_1 \oplus \cdots \oplus A_r$, for some $A_1, \ldots, A_r \in \mathcal{A}$ (and some r).

2) r and $A_1, \ldots, A_r$ are unique (up to a reordering).

3) Suppose A' is similar to $A'_1 \oplus \cdots \oplus A'_s$, with $A'_1, \ldots, A'_s \in \mathcal{A}$. Then the matrices A and A' are similar if and only if $r = s$ and, after a reordering, we have $A_i = A'_i$ for every i.

Specific Canonical Forms

- **Jordan Canonical Form** is the main topic of Section 10.2.
- **Rational Canonical Form** is described in the following exercise.

(10.1.5) **Exercise** (Rational Canonical Form)**.** For $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 \in F[x]$, we use A_f to denote the *companion matrix* of $f(x)$, which means

$$A_f = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ 0 & 0 & 1 & \cdots & 0 & -a_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -a_{n-1} \end{bmatrix}.$$

Show:

1) The minimal polynomial and characteristic polynomial of A_f are $f(x)$.

2) The $F[x]$-module corresponding to A_f is isomorphic to $F[x]/\langle f(x)\rangle$.
[*Hint:* Exercise 6.3.5(7).]

3) Any $n \times n$ matrix A over F is similar to a matrix of the form
$$A_{f_1} \oplus A_{f_2} \oplus \cdots \oplus A_{f_k},$$
where each f_i is a power of a monic irreducible polynomial. Furthermore, the choice of $f_1, f_2, \ldots, f_k$ is unique (up to a permutation).

4) The same as the preceding, except we change the assumption on f_i to:
f_i is a divisor of f_{i+1} for every i (and no f_i is a unit). However, in this case, the choice of $f_1, f_2, \ldots, f_k$ is entirely unique (not merely up to permutations).

§10.2. Jordan Canonical Form

(10.2.1) **Assumption.** In the remainder of this chapter:

All vector spaces are over $\mathbb{C}$. (That is, we assume $F = \mathbb{C}$.)

We make this assumption because it implies that every polynomial in $F[x]$ is a product of linear factors (so Proposition 8.2.12 tells us that every square matrix is similar to an upper-triangular matrix).

It is important to realize that some matrices cannot be diagonalized (i.e., they are not similar to a diagonal matrix), even over $\mathbb{C}$:

(10.2.2) **Exercise.** Let $J = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. Show:

1) The minimal polynomial of J is x^4.
   [*Hint:* Verify that $J^4 = 0$, but $J^3 \neq 0$.]
2) J is not similar to a diagonal matrix (over any field).
   [*Hint:* Proposition 8.3.7 or Exercise 8.3.8.]

Jordan Canonical Form brings a matrix as close to being diagonal as possible. Namely, as in Exercise 10.2.2, the only nonzero entries that are not on the main diagonal are required to be right next to it. Furthermore, these nonzero entries off the main diagonal are required to be 1:

(10.2.3) **Notation.** For $\lambda \in \mathbb{C}$ and $k \in \mathbb{N}^+$, let

$$J_k(\lambda) = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \mathbf{0} \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ \mathbf{0} & & & & \lambda \end{bmatrix} \qquad \text{be defined by} \qquad J_k(\lambda)_{i,j} = \begin{cases} \lambda & \text{if } i = j, \\ 1 & \text{if } i+1 = j, \\ 0 & \text{otherwise.} \end{cases}$$

In other words, $J_k(\lambda)$ is the $k \times k$ matrix in which in all entries are 0, except:

- every entry on the main diagonal is λ, and
- every entry in the diagonal just above the main diagonal is 1.

$J_k(\lambda)$ is often called a "***Jordan block***."

(10.2.4) **Example.** The matrix J of Exercise 10.2.2 is $J_4(0)$. Also:

$$J_5(2) = \begin{bmatrix} \mathbf{2} & \mathbf{1} & 0 & 0 & 0 \\ 0 & \mathbf{2} & \mathbf{1} & 0 & 0 \\ 0 & 0 & \mathbf{2} & \mathbf{1} & 0 \\ 0 & 0 & 0 & \mathbf{2} & \mathbf{1} \\ 0 & 0 & 0 & 0 & \mathbf{2} \end{bmatrix}, \qquad J_4(-3) = \begin{bmatrix} \mathbf{-3} & \mathbf{1} & 0 & 0 \\ 0 & \mathbf{-3} & \mathbf{1} & 0 \\ 0 & 0 & \mathbf{-3} & \mathbf{1} \\ 0 & 0 & 0 & \mathbf{-3} \end{bmatrix}.$$

(10.2.5) **Exercise.**

1) What is the characteristic polynomial of $J_k(\lambda)$?
   [*Hint:* $J_k(\lambda)$ is upper-triangular, with λ's on the diagonal (or $J_k(\lambda) = \lambda I + J_k(0)$, and $J_k(0)^k = 0$).]
2) What is the minimal polynomial of $J_k(\lambda)$?
3) Is $J_k(\lambda)$ diagonalizable?
   [*Hint:* The answer is different for $k = 1$ than for $k > 1$.]

(10.2.6) **Proposition.** *If $A = J_k(a)$, for some $a \in \mathbb{C}$ and $k \in \mathbb{N}^+$, then the corresponding $\mathbb{C}[x]$-module M_A is cyclic, torsion, and primary.*

Proof. (cyclic) For simplicity, we first consider the case where $a = 0$, which means $A = J_k(0)$. Let $\{e_1, \ldots, e_k\}$ be the standard basis of $\mathbb{C}^k$, and note that the columns of $J_k(0)$ are (from left to right):

$$0,\ e_1,\ e_2, \ldots,\ e_{k-1},$$

so the ith column of $J_k(0)$ is e_{i-1}, for $i \geq 2$. This means

$$Ae_i = J_k(0)e_i = e_{i-1} \quad \text{for } 2 \leq i \leq k.$$

Hence, for $v = e_k$, we see, by induction on i, that $A^i v = e_{k-i}$ for $0 \leq i < k$. Hence

$$\mathbb{C}[x]\, v \supset \langle\, A^i v \mid 0 \leq i < k \,\rangle = \langle\, e_k, e_{k-1}, e_{k-2}, \ldots, e_2, e_1 \,\rangle = \mathbb{C}^k,$$

since $\{e_1, \ldots, e_k\}$ is a basis. So M_A is cyclic.

Essentially the same proof applies in the general case, because

$$\langle\, J_k(0)^i\, v \mid 0 \leq i < k \,\rangle \subseteq \mathbb{C}[x]\, v,$$

since $J_k(0)^i = (A - aI)^i \in \mathbb{C}[A]$.

(torsion and primary) Let $f(x) = (x-a)^k$, so

$$f(A) = (A - aI)^k = \big(J_k(a) - aI\big)^k = J_k(0)^k = 0.$$

Then $f(x)$ annihilates M_A (since, for any $m \in M_A$, we have $f(x)\,m = f(A)\,m = 0\,m = 0$). Since $f(x)$ is a power of the irreducible polynomial $x-a$, this implies that M_A is torsion and primary. □

(10.2.7) **Exercise.** Conversely, if M is a nonzero, cyclic, primary, torsion module over $\mathbb{C}[x]$, then $M \cong M_{J_k(a)}$, for some $a \in \mathbb{C}$ and $k \in \mathbb{N}^+$.

[*Hint:* By assumption, M is cyclic, and its annihilator is $\mathbb{C}[x]\,\pi(x)^k$, for some irreducible polynomial $\pi(x) \in \mathbb{C}[x]$, and some $k \in \mathbb{N}^+$. The Fundamental Theorem of Algebra (8.2.14) tells us that $p(x)$ has a root a in $\mathbb{C}$, which means that $p(x)$ is divisible by the linear polynomial $x-a$. So irreducibility implies that $\pi(x)$ is $x-a$ (times a unit). Then the cyclic modules M and $M_{J_k(a)}$ have the same annihilator, so they are isomorphic.]

(10.2.8) **Proposition.** *Suppose $k, \ell \in \mathbb{N}^+$ and $a, b \in \mathbb{C}$. If $M_{J_k(a)} \cong M_{J_\ell(b)}$, then $k = \ell$ and $a = b$.*

Proof. We know that $J_k(a)$ is similar to $J_\ell(b)$ (see Exercise 10.1.2). Since $J_k(a)$ is a $k \times k$ matrix and $J_\ell(b)$ is an $\ell \times \ell$ matrix, this implies $k = \ell$. It also implies that $J_k(a)$ and $J_\ell(b)$ have the same eigenvalues (cf. Remark 8.3.9). Since a is the only eigenvalue of $J_k(a)$, and b is the only root of eigenvalue of $J_\ell(b)$, we conclude that $a = b$. □

Combining Propositions 10.2.6 and 10.2.8 and Exercise 10.2.7 shows that the set $\mathcal{J}$ of Jordan blocks $J_k(\lambda)$ satisfies the hypotheses of Exercise 10.1.4. This yields the following conclusion:

(10.2.9) **Theorem** (Jordan Canonical Form)**.** *Let $A \in \mathrm{Mat}_{n\times n}(\mathbb{C})$.*

1) *A is similar to a matrix of the form $J_{k_1}(\lambda_1) \oplus \cdots \oplus J_{k_r}(\lambda_r)$, for some r, $k_1, \ldots, k_r \in \mathbb{N}^+$ and $\lambda_1, \ldots, \lambda_r \in \mathbb{C}$. This is called a **Jordan Canonical Form** of A.*
2) *Furthermore, the blocks in a Jordan Canonical Form are uniquely determined (up to a reordering).*

(10.2.10) **Corollary.** *Let $A, A' \in \mathrm{Mat}_{n\times n}(\mathbb{C})$. Suppose A is similar to $J_1 \oplus \cdots \oplus J_r$ and A' is similar to $J'_1 \oplus \cdots \oplus J'_s$, where each J_i and J'_i is a Jordan block. Then the matrices A and A' are similar if and only if $r = s$ and, after a reordering, we have $J_i = J'_i$ for every i.*

(10.2.11) **Example.** Here is a matrix in Jordan Canonical Form:

$$J_3(3) \oplus J_2(3) \oplus J_2(6) = \begin{bmatrix} \boxed{\begin{matrix} 3 & 1 & \\ & 3 & 1 \\ & & 3 \end{matrix}} & & \\ & \boxed{\begin{matrix} 3 & 1 \\ & 3 \end{matrix}} & \\ & & \boxed{\begin{matrix} 6 & 1 \\ & 6 \end{matrix}} \end{bmatrix} = \begin{bmatrix} \mathbf{3} & \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{3} & \mathbf{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{3} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{3} & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{6} & \mathbf{1} \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{6} \end{bmatrix}.$$

(10.2.12) *Remark.* A square matrix $A = (a_{i,j})$ is in Jordan Canonical Form if and only if

- every nonzero entry is on either the main diagonal or immediately above the main diagonal,
- every nonzero entry immediately above the diagonal is a 1, and
- whenever a 1 appears above the diagonal, the two diagonal entries adjacent to it (the one to its left and the one below it) are equal.

More precisely, for $1 \le i, j \le n$, we have:

$$a_{i,j} \neq 0 \Rightarrow j \in \{i, i+1\}, \quad a_{i,i+1} \in \{0, 1\}, \quad \text{and} \quad a_{i,i+1} = 1 \Rightarrow a_{i,i} = a_{i+1,i+1}.$$

(10.2.13) *Remark.* Although we have only discussed matrices whose entries are complex numbers, the same ideas apply more generally. Namely, suppose $A \in \mathrm{Mat}_{n\times n}(F)$, where F is any field. Then A is similar to a matrix in Jordan Canonical Form if and only if every eigenvalue of A is in F. (In other words, if and only if the characteristic polynomial of A factors as a product of linear polynomials in $F[x]$.)

(10.2.14) **Observation.** *It is easy to find the eigenvalues, characteristic polynomial, and minimal polynomial of any matrix in Jordan Canonical Form.*

1) *Since the Jordan Canonical Form of any matrix is upper triangular, its eigenvalues are the numbers that appear on the main diagonal* (*see Corollary 8.2.10*). *For example, the eigenvalues of the matrix in Example 10.2.11 are* 3 *and* 6.
2) *In general, the characteristic polynomial of a block-diagonal matrix is the product of the characteristic polynomials of the blocks* (*cf. Proposition 8.2.6*). *Since the characteristic polynomial of* $J_k(\lambda)$ *is* $(x-\lambda)^k$ (*see Exercise 10.2.5*(1)), *we conclude that*
$$\text{the characteristic polynomial of } \bigoplus_{i=1}^{r} J_{k_i}(\lambda_i) \text{ is } \prod_{i=1}^{r}(x-\lambda_i)^{k_i}.$$
3) *On the other hand, the minimal polynomial of a block-diagonal matrix is the* ***least common multiple*** (*not the product!*) *of the minimal polynomials of the blocks* (*see Exercise 10.2.16*). *Therefore, if the set of eigenvalues is* $\{\mu_1,\dots,\mu_s\}$ *where the* μ_i*'s are* distinct, *and we let* $m_j = \max\{\, k_i \mid \lambda_i = \mu_j \,\}$ *for each* j, *then, since* $\operatorname{lcm}((x-\mu_j)^k,(x-\mu_j)^\ell) = (x-\mu_j)^{\max(k,\ell)}$,
$$\text{the minimal polynomial of } \bigoplus_{i=1}^{r} J_{k_i}(\lambda_i) \text{ is } \prod_{j=1}^{s}(x-\mu_j)^{m_j}.$$

(10.2.15) *Remark.* The minimal polynomial is *not* equal to the characteristic polynomial, unless the occurrences of each eigenvalue λ are all in one block. (That is, $\lambda_i \neq \lambda_j$ whenever $i \neq j$.) This is because the exponent of $(x-\lambda)$ in the characteristic polynomial is the *sum* of the sizes of the λ-blocks, but the exponent in the minimal polynomial is the *maximum* of the sizes.

(10.2.16) **Exercise.** Assume $m_i(x)$ is the minimal polynomial of A_i, for $i = 1, 2, \dots, n$. Show that the minimal polynomial of $A_1 \oplus A_2 \oplus \cdots \oplus A_n$ is $\operatorname{lcm}\big(m_1(x), m_2(x), \dots, m_n(x)\big)$.
[*Hint:* $p\big(A_1 \oplus \cdots \oplus A_n\big) = p(A_1) \oplus \cdots \oplus p(A_n)$ for any $p(x) \in F[x]$.]

(10.2.17) **Corollary** (Cayley-Hamilton Theorem)**.** *Every square matrix satisfies its characteristic polynomial. More precisely, if* $f(x)$ *is the characteristic polynomial of a matrix* $A \in \mathrm{Mat}_{n\times n}(F)$, *then* $f(A) = 0$.

Proof. Observation 10.2.14 shows that the characteristic polynomial is always a multiple of the minimal polynomial (because the product of any number of polynomials is always a multiple of the least common multiple, or, if you prefer to use Remark 10.2.15, it is because the sum of any collection of natural numbers is always at least as large as the maximum of the numbers). □

(10.2.18) **Warning.** Students sometimes erroneously believe the Cayley-Hamilton Theorem is obvious, thinking that, since $p(x) = \det(xI - A)$, we can simply substitute A for x, to obtain
$$p(A) = \det(AI - A) = \det(A - A) = 0.$$
However, this argument is nonsensical. For example, the expression $p(A)$ on the left-hand side is an $n \times n$ matrix, but the number 0 on the right-hand side is a scalar (since it is the *determinant* of a matrix, not the matrix itself), so the two sides certainly cannot be equal (unless $n = 1$).

With Observation 10.2.14 or Remark 10.2.15 in mind, it is not difficult to find all of the $n \times n$ matrices with a given characteristic polynomial or minimal polynomial (or both), up to similarity. It is much like finding all of the abelian groups of a given finite order, up to isomorphism (as in Exercise 7.3.7).

(10.2.19) **Example.** Jordan Canonical Form allows us to find (up to similarity) all of the matrices in $\mathrm{Mat}_{5\times5}(\mathbb{C})$ whose characteristic polynomial is $p(x) = (x-4)^2(x+9)^3$. First, note that the eigenvalues of the matrix must be 4 and -9.

- The exponent of $x-4$ in $p(x)$ is 2. Since the partitions of 2 are 2 and 1 + 1, this implies that the configuration of the 4-blocks in the Jordan Canonical Form is either $J_2(4)$ or $J_1(4) \oplus J_1(4)$.

- The exponent of $x+9$ in $p(x)$ is 3. Since the partitions of 3 are 3, 2 + 1, and 1 + 1 + 1, this implies that the configuration of the (-9)-blocks in the Jordan Canonical Form is either $J_3(-9)$, or $J_2(-9) \oplus J_1(-9)$, or $J_1(-9) \oplus J_1(-9) \oplus J_1(-9)$.

Combining each possible configuration of 4-blocks with each possible configuration of (-9)-blocks yields:

$$J_2(4) \oplus J_3(-9), \qquad J_1(4) \oplus J_1(4) \oplus J_3(-9),$$
$$J_2(4) \oplus J_2(-9) \oplus J_1(-9), \qquad J_1(4) \oplus J_1(4) \oplus J_2(-9) \oplus J_1(-9),$$
$$J_2(4) \oplus J_1(-9) \oplus J_1(-9) \oplus J_1(-9), \quad J_1(4) \oplus J_1(4) \oplus J_1(-9) \oplus J_1(-9) \oplus J_1(-9).$$

Therefore, up to similarity, the only 5×5 matrices with this characteristic polynomial are:

$$\begin{bmatrix} \mathbf{4} & \mathbf{1} & 0 & 0 & 0 \\ 0 & \mathbf{4} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{-9} & \mathbf{1} & 0 \\ 0 & 0 & 0 & \mathbf{-9} & \mathbf{1} \\ 0 & 0 & 0 & 0 & \mathbf{-9} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{4} & 0 & 0 & 0 & 0 \\ 0 & \mathbf{4} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{-9} & \mathbf{1} & 0 \\ 0 & 0 & 0 & \mathbf{-9} & \mathbf{1} \\ 0 & 0 & 0 & 0 & \mathbf{-9} \end{bmatrix},$$

$$\begin{bmatrix} \mathbf{4} & \mathbf{1} & 0 & 0 & 0 \\ 0 & \mathbf{4} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{-9} & \mathbf{1} & 0 \\ 0 & 0 & 0 & \mathbf{-9} & 0 \\ 0 & 0 & 0 & 0 & \mathbf{-9} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{4} & 0 & 0 & 0 & 0 \\ 0 & \mathbf{4} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{-9} & \mathbf{1} & 0 \\ 0 & 0 & 0 & \mathbf{-9} & 0 \\ 0 & 0 & 0 & 0 & \mathbf{-9} \end{bmatrix},$$

$$\begin{bmatrix} \mathbf{4} & \mathbf{1} & 0 & 0 & 0 \\ 0 & \mathbf{4} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{-9} & 0 & 0 \\ 0 & 0 & 0 & \mathbf{-9} & 0 \\ 0 & 0 & 0 & 0 & \mathbf{-9} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{4} & 0 & 0 & 0 & 0 \\ 0 & \mathbf{4} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{-9} & 0 & 0 \\ 0 & 0 & 0 & \mathbf{-9} & 0 \\ 0 & 0 & 0 & 0 & \mathbf{-9} \end{bmatrix}.$$

(10.2.20) **Example.** Assume $r, s \in \mathbb{C}$ (and $r \neq s$). We can find (up to similarity) all of the 7×7 matrices whose characteristic polynomial is $p(x) = (x-r)^2(x-s)^5$, and whose minimal polynomial is $m(x) = (x-r)^2(x-s)^2$. First, note that the eigenvalues of the matrix are r and s.

- The exponent of $x-r$ in the characteristic polynomial $p(x)$ is 2, so the sizes of the r-blocks in the Jordan Canonical Form must add up to 2. However, the exponent of $x-r$ in the minimal polynomial $m(x)$ is 2, so the largest r-block is of size 2. Hence, $J_2(r)$ is the only r-block.
- The exponent of $x-s$ in $p(x)$ is 5, so the sizes of the s-blocks must add up to 5. Also, the exponent of $x-s$ in $m(x)$ is 2, so the size of the largest s-block is 2. The partitions of 5 in which the largest summand is 2 are: 2 + 2 + 1 and 2 + 1 + 1 + 1, so this implies that the configuration of the s-blocks is either $J_2(s) \oplus J_2(s) \oplus J_1(s)$ or $J_2(s) \oplus J_1(s) \oplus J_1(s) \oplus J_1(s)$.

By combining the r-block $J_2(r)$ with each of the two possible configurations of s-blocks, we see that (up to similarity) the only 7×7 matrices with the given characteristic polynomial and minimal polynomial are:

$$\begin{bmatrix} \boldsymbol{r} & \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \boldsymbol{r} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \boldsymbol{s} & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \boldsymbol{s} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \boldsymbol{s} & \mathbf{1} & 0 \\ 0 & 0 & 0 & 0 & 0 & \boldsymbol{s} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \boldsymbol{s} \end{bmatrix}, \quad \begin{bmatrix} \boldsymbol{r} & \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \boldsymbol{r} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \boldsymbol{s} & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \boldsymbol{s} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \boldsymbol{s} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \boldsymbol{s} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \boldsymbol{s} \end{bmatrix}.$$

(10.2.21) **Exercise.** Assume $a, b \in \mathbb{C}$, and $a \neq b$. Find (up to similarity over $\mathbb{C}$):

1) all of the 5×5 matrices whose characteristic polynomial is $(x-a)^5$,
2) all of the 5×5 matrices whose minimal polynomial is $(x-a)^2$,

3) all of the 5×5 matrices whose characteristic polynomial is $(x-a)(x-b)^4$,
4) all of the 6×6 matrices whose minimal polynomial is $(x-a)^2(x-b)^2$.
5) all of the 6×6 matrices whose minimal polynomial is $(x-a)^3(x-b)^2$.
6) all of the matrices A, such that
   - the characteristic polynomial of A is $(x-a)^5(x-b)^3$, and
   - the minimal polynomial of A is $(x-a)^3(x-b)^2$.

Index