Algebra

Justin Lawrence

May 29, 2024

Preface

This book is a bit of an experiment. It started as deciding to put together some notes on everything I learned in undergraduate algebra and has slowly expanded to become something more : an attempt to take a more systmatic approach to algebra as a whole. Some parts are more in flux than others, the foundations section in particular will probably go through many rewrites.

To use the fun buzzwords that every textbook has, *a certain level of mathematical maturity is assumed*. Also assumed is an understanding of linear algebra at an introductory undergraduate level. No familiarity with categories or abstract algebra is assumed. The book may be better titled Intermediate Algebra, as its level of sophistication lies somewhere between an undergraduate and graduate approach to algebra. Sections marked with $*$ are considered optional : it is fairly safe to assume that no future sections will depend on them.

As with most mathematics textbooks, none of the ideas contained within it were mine to begin with. Much of the book is based off of Serge Lang's Algebra [Lan05], a wonderful reference text that I suggest you pick up yourself. Other inspirations include Nathan Jacobson's Basic Algebra I [Jac09] for the earlier sections (although I do somewhat hate that book) and Steven Rotman's Advanced Linear Algebra [Rot07]. I am of course also in debt to the many professors I've had so far in Algebra throughout my years of education. In particular, it was Vinyak Vatsal who originally taught me the basics of algebra (from that cursed Jacobson book), and Lior Silberman who taught me to see algebra in a more unified way. The linear algebra section of this book in particular is essentially a synthesis of/selection of topics from a course in advanced linear algebra he's taught a few times at UBC, which I highly recommend taking if you get the chance.

Finally, I would like to thank two other people. My friend Boris for doing his own version of this for forcing me to stop procrastinating on this project. And second I would like to thank Ben Williams, both for his wonderful insights and advice over the past year, and for making me install a spell checker. I'm sure there are still grammatical errors in this text that would cause you to lose your mind slightly, but please know that I really did try to fix them[1].

[1]Somehow, despite only being fluent in one language, I'm quite terrible at it.

Contents

Chapter 1

Foundations

Unfortunately, we cannot immediately jump into algebra without first having a strong understanding of some basic mathematical concepts. The first of these is a combination of basic facts about prime numbers, factorization, and equivalence relations. These results will be used immediately and frequently in our study of algebra, and you should have a firm understanding of them before moving on. The second is an overview of Zorn's lemma. This will not be used until a fair bit into our studies, and can be skipped for now if desired. When you come back to it, it is not essentially that you understand the proof of the lemma, just how to apply it.[1]

1.1 Primes and Equivalence Classes

This section collects results from a similar one in [Jac09] (which is honestly one of his better expository moments), and is here for your convenience. If you are not familiar with these concepts to some degree already, please read the corresponding sections in [Jac09]. We start with some very basic definitions.

Definition 1.1.1. Let $a, b \in \mathbb{Z}$. We say that a divides b, or a is a divisor of b, denoted $a \mid b$, if there exists some $x \in \mathbb{Z}$ such that $b = az$. We say that a number $p \in \mathbb{Z}$ is prime if its only divisors are $\pm 1, \pm p$. By convention, we do not consider ± 1 to be prime.

Note that if $b \mid c$ and $a \mid b$, then $a \mid c$. Indeed, we use this fact to prove one of the most fundamental theorems of mathematics.

Theorem 1.1.2 (Prime Factorization)**.** *Any number $n \in \mathbb{N}$ has a unique (up to order of primes) representation in the form*

$$n = p_1^{e_1} \cdots p_r^{e_r} \tag{1.1}$$

where $p_i \in \mathbb{N}$ are prime, and $e_i \in \mathbb{N}$.

For a proof, see [Jac09] (and apply this statement to every part of this section).

[1] It is a good exercise in dealing with abstract concepts to understand it however.

Definition 1.1.3. Let $a, b \in \mathbb{Z}\backslash\{0\}$. Then we define

1. A greatest common denominator GCD of a, b to be any number $c \in \mathbb{Z}$ such that $c \mid a, b$ and if $d \mid a, b$ then $d \mid c$.
2. A least common multiple of a, b to be any number $c \in \mathbb{Z}$ such that if $a, b \mid c$ and if $a, b \mid d$ then $c \mid d$.

Theorem 1.1.4. *Let $a, b \in \mathbb{Z}\backslash\{0\}$. Let $|a| = p_1^{e_1} \cdots p_r^{e_r}, |b| = p_1^{f_1} \cdots p_r^{f_r}$ be prime factorizations, where we allow $e_i, f_i = 0$. Set $M_i = \max(e_i, f_i), m_i = \min(e_i, f_i)$. Then*

1. *The GCDs of a, b are $\pm p_1^{m_1} \cdots p_r^{m_r}$.*
2. *The LCMs of a, b are $\pm p_1^{M_1} \cdots p_r^{M_r}$.*

Because of the above theorem, we usually denote the positive GCD/LCM by $(a, b), [a, b]$.

Proposition 1.1.5. *Let $a, b \in \mathbb{N}$. Then*

$$(a, b)[a, b] = ab$$

Theorem 1.1.6 (Division Algorithm)**.** *Suppose $a, b \in \mathbb{Z}$, with $b \neq 0$. Then there exists some $q, r \in \mathbb{Z}$ such that $0 \leq r < |b|$ and*

$$a = qb + r$$

Corollary 1.1.6.1 (Bézout's identity)**.** *Suppose $a, b \in \mathbb{Z}$ are non-zero. Then there exist $m, n \in \mathbb{Z}$ such that $ma + nb = (a, b)$ and*

$$\{ca + db \mid c, d \in \mathbb{Z}\} = \{c(a, b) \mid c \in \mathbb{Z}\}$$

Now, we move on to equivalence relations.

Definition 1.1.7. Let S be a set. A relation R on S is a subset of $S \times S$. If $(x, y) \in R$, we denote this by xRy.

Definition 1.1.8. Let S be a set. An equivalence relation $\sim$ on S is a relation satisfying the following three axioms for all $x, y, z \in S$

1. $x \sim x$ (Symmetry)
2. $x \sim y \Rightarrow y \sim x$ (Reflexivity)
3. $x \sim y, y \sim z \Rightarrow x \sim z$ (Transitivity)

Equivalence relations allow us to partition sets into what are called equivalence classes.

Definition 1.1.9. Let $x \in S$, and $\sim$ be an equivalence relation on S. The equivalence class of x, denoted $[x]_\sim$ or just $[x]$, is defined in the following manner.

$$[x] = \{y \in S \mid x \sim y\}$$

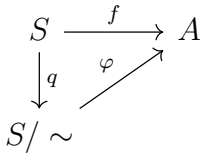
Proposition 1.1.10. *$[x] = [y]$ if and only if $x \sim y$. The set $S/\sim$ of equivalence classes in S (called the quotient set of S) is therefore a partition of S, that is a division of S into disjoint subsets whose union is S.*

The map $q : S \to S/\sim$ defined by $x \mapsto [x]$ is called the quotient function. One can actually go the other way as well.

Proposition 1.1.11. *Let $\pi \subset \mathcal{P}(S)$ be a partition. Then there exists a unique equivalence relation $\sim$ on S such that $S/\sim = \pi$. In particular, this $\sim$ is defined by two elements being equivalent if and only if they lie in the same set in the partition.*

Finally, sufficiently well-behaved functions on S will induce unique maps on the quotient set.

Theorem 1.1.12. *Suppose $f : S \to A$ is a mapping between sets, and $\sim$ is an equivalent relation on S with quotient function q. If f is such that $x \sim y \Rightarrow f(x) = f(y)$, for all $x, y \in S$, then there exists a unique function $\varphi : S/\sim \to A$ such that $\varphi \circ q = f$, that is such that the following diagram commutes.*



1.2 Zorn's Lemma

This section is a streamlined version of a similar section in [Lan05].

Zorn's lemma is to algebra what Fourier transforms are to physics. Nobody will every explicitly teach it to you, but it gets brought up constantly and at some point you seem to be expected to just learn it via osmosis. I open with it in hopes that, if you haven't learned it before, now will be your chance to learn about Zorn's lemma and its proof.

We begin with a definition.

Definition 1.2.1. Let S be a set. A partial ordering of S is a relation $\leq$ between elements of S satisfying the following axioms $\forall x, y, z \in S$

1. $x \leq x$
2. $x \leq y \wedge y \leq z \Rightarrow x \leq z$
3. $x \leq y \wedge y \leq x \Rightarrow x = y$

Note. We do not require that every pair of elements in S be comparable. If this additional condition is satisfied, we call $\leq$ a *total ordering*, and say that S is *totally ordered*. Totally ordered subsets of partially ordered sets are often called chains. If $x \leq y$ and $x \neq y$, we write that $x < y$.

We follow this up with a collection of definitions related to Definition 1.2.1.

Definition 1.2.2. Let S be an ordered set. A smallest element of S is an element $a \in S$ such that $a \leq x$ for all $x \in S$, with a greatest element defined similarly. A maximal element $m \in S$ is an element such that if $m \leq x \Rightarrow x = m$, and a minimal element is defined similarly.

Note. Maximal and greatest elements are not identical notions. Indeed, one can note that maximal elements need not be a greatest element, and that greatest elements are unique when they exist (while maximal elements may not be). A similar result hold minimal and smallest elements.

Definition 1.2.3. Let $T \subseteq S$ be a subset of a partially ordered set. An upper bound of T in S is an element $a \in S$ such that $t \leq a$ for all $t \in T$. A least upper bound of T in S is an upper bound $b \in S$ such that for any other upper bound $a \in S$, $b \leq a$. A set is inductively ordered if every non-empty totally ordered subset has an upper bound, and strictly so if it has a least upper bound.

Definition 1.2.4. Let A be a non-empty partially and strictly inductively ordered set. A map $f : A \to A$ is increasing if, for all $x \in S$, $x \leq f(x)$.

Definition 1.2.5. Let A be a non-empty partially and strictly inductively ordered set, and $f : A \to A$ an increasing map. Pick some $a \in A$, and let $B \subseteq A$. We say that B is admissible with respect to a if

1. $a \in B$
2. $f(B) \subseteq B$
3. Whenever T is a non-empty totally ordered subset of B, the least upper bound of T in A lies in B

Definition 1.2.6. Let A be a non-empty partially and strictly inductively ordered set with minimal element $a \in A$, and $f : A \to A$ an increasing map. We define $M(A, f)$ to be the intersection of all admissible subsets of A with respect to a.

Note. It is not too difficult to see that $M(A, f)$ is the smallest admissible subset of A with respect to a, and is contained in all admissible subsets of A with respect to a. Furthermore, $M(A, f)$ is strictly inductively ordered.

Definition 1.2.7. Let A be a non-empty partially and strictly inductively ordered set with minimal element $a \in A$, and $f : A \to A$ an increasing map. We say that $c \in M(A, f)$ is an *extreme point* of $M(A, f)$ if $x \in M(A, f), x < c \Rightarrow f(x) \leq c$. We further define for such points that

$$M_c(A, f) = \{x \in M(A, f) \mid x \leq c \vee f(c) \leq x\}$$

Note. $a \in M_c(A, f)$, so this set is necessarily non-empty.

We next build up a series of lemmas related to these definitions. The point of this is to prove the third of the lemmas, which will be used to prove a subsequent theorem.

Lemma 1.2.8. *Let A be a non-empty partially and strictly inductively ordered set with minimal element $a \in A$, $f : A \to A$ an increasing map, and $c \in M(A, f)$ be an extreme point. Then $M_c(A, f) = M(A, f)$.*

Proof. It suffices to prove that $M_c(A, f)$ is admissible with respect to a. We already have $a \in M_c(A, f)$. Suppose $x \in M_c(A, f)$. If $x < c$, then since c is an extreme point we get $f(x) \leq c$, so $f(x) \in M_c(A, f)$. If $x = c$, then $f(c) \leq f(x) \Rightarrow f(x) \in M_c(A, f)$. If $f(c) \leq x$, then $f(c) \leq x \leq f(x) \Rightarrow f(c) \in M_c(A, f)$. Thus, $f(M_c(A, f)) \subseteq M_c(A, f)$ as required. Let $T \subseteq M_c(A, f)$ be a non-empty totally ordered subset, and $b \in M(A, f)$ the least upper bound of T in $M(A, f)$. Pick any $x \in T$. If $f(c) \leq x$, then $f(c) \leq b$ so $b \in M_c(A, f)$. If $x \leq c$ for all $x \in T$, then $b \leq c \Rightarrow b \in M_c(A, f)$, as required. □

Lemma 1.2.9. *Let A be a non-empty partially and strictly inductively ordered set with minimal element $a \in A$, and $f : A \to A$ an increasing map. Then every element of $M(A, f)$ is an extreme point.*

Proof. Let $E \subseteq M(A, f)$ be the set of all extreme points. Again, it suffices to show that E is admissible with respect to a. a is vacuously extreme, so $a \in E$. Now, pick any $c \in E$, $x \in M(A, f)$. Suppose $x < f(c)$. By lemma 1.2.8, $M_c(A, f) = M(A, f)$, so $x \leq c$ or $f(c) \leq x$. If $x < c$, then $f(x) \leq f(c)$. If $x = c$, then $f(x) \leq f(c)$. This proves that $f(c) \in E$ as desired. Finally, let $T \subseteq E$ be a non-empty totally ordered subset, and $b \in M(A, f)$ the least upper bound of T in $M(A, f)$. Suppose $x \in M(A, f)$ and $x < b$. Then $\exists c \in T$ such that $x \leq c$ (indeed, we otherwise get by lemma 1.2.8 that $f(c) \leq x$ for all $c \in T$, and hence $c \leq x$ for all $c \in T$, so $b \leq x$). If $x < c$, then $f(x) \leq c \leq b$ so $f(x) \leq b$. If $x = c$, then by lemma 1.2.8 we must get $f(x) \leq b$ (as otherwise $b \leq x$). This shows that $b \in E$, and hence completes the proof. □

Lemma 1.2.10. *Let A be a non-empty partially and strictly inductively ordered set with minimal element $a \in A$, and $f : A \to A$ an increasing map. Then $M(A, f)$ is totally ordered.*

Proof. Pick any $x, y \in M(A, f)$. By lemma 1.2.9, y is an extreme point of $M(A, f)$, so by lemma 1.2.8 either $x \leq y$ or $y \leq f(y) \leq x \Rightarrow y \leq x$. □

Using this result, we prove a powerful theorem of which Zorn's lemma is a corollary.

Theorem 1.2.11 (Bourbaki's Theorem)**.** *Let A be a non-empty partially and strictly inductively ordered set, and $f : A \to A$ an increasing map. Then $\exists x_0 \in A$ such that $f(x_0) = x_0$, that is f has a fixed point.*

Proof. Suppose that A is totally ordered. Then since it has a least upper bound $b \in A$, $b \leq f(b) \leq b \Rightarrow b = f(b)$, as required. Otherwise, it suffices to find an admissible totally ordered subset of A. Pick some $a \in A$ and let B be the set of elements $x \in A$ such that $x < a$. Then $A \backslash B$ is admissible with respect to a, and a is a minimal element of $A \backslash B$, so we may assume without loss of generality that A has a minimal element $a \in A$. By lemma 1.2.10, $M(A, f)$ is the desired totally ordered admissible subset. □

Corollary 1.2.11.1 (Weak Zorn's Lemma)**.** *Let A be a non-empty partially and strictly inductively ordered set. Then A has a maximal element.*

Proof. Suppose not. Then for any $x \in A$, there exists some $y_x \in A$ such that $x < y_x$[2], as otherwise x would be maximal. Let $f : A \to A$ be defined by $f : x \mapsto y_x$. Then f is increasing, so by Theorem 1.2.11 f has a fixed point, which is impossible. □

Corollary 1.2.11.2 (Zorn's Lemma)**.** *Let A be a non-empty partially and inductively ordered set. Then A has a maximal element.*

Proof. Let B be the set of non-empty totally ordered subsets of A. Then B is not empty, as any singleton is totally ordered. If $X, Y \in B$, we define a partial order $\leq$ on B by $X \leq Y \iff X \subseteq Y$. In fact, this makes B strictly inductively ordered. To see this, let $T = \{X_i\}_{i \in I} \subset B$ be totally ordered, and let $Z = \cup_{i \in I} X_i$. Pick any $x, y \in Z$. Then $x \in X_i, y \in Y_j$ for some $i, j \in I$. Since T is totally ordered, we get (without loss of generality) $X_i \subseteq X_j$, so since $X_j \in B$ we must have $x \leq y$ or $y \leq x$. Thus, Z is totally ordered, and hence clearly a least upper bound of T. Therefore, B is a non-empty partially and strictly inductively ordered set, and therefore has a maximal element $X_0 \in B$. Since A is inductively ordered, X_0 has an upper bound $m \in A$. We'll show that m is a maximal element of S. Indeed, suppose that $x \in S$ and $m \leq x$. Then $X_0 \cup \{x\}$ is totally ordered, so by the maximality of X_0 we must get $X_0 = X_0 \cup \{x\} \Rightarrow x \in X_0 \Rightarrow x \leq m$, so $x = m$, as was to be shown. □

Note. The non-empty condition comes from the definition of an inductively ordered set, and isn't really needed. Indeed, suppose that A is a non-empty partially ordered set such that every totally ordered subset has an upper bound. Then in particular every non-empty totally ordered subset has an upper bound, so by Zorn's lemma A has a maximal element.

Zorn's lemma turns out to be equivalent to the axiom of choice, but as this is not a book on set theory we won't get further into that here. If you're interested in learning more about that, it may be worth starting at this rabbit hold of a Wikipedia page[3].

[2]This choice of y_x is invoking the axiom of choice

[3]`https://en.wikipedia.org/wiki/Axiom_of_choice#Equivalents`

Part I

Basic Algebra

Chapter 2

Groups

2.1 Basic Definitions

We start with the basic definitions of group theory.

Definition 2.1.1. A monoid is a tuple $(M, 1, \cdot)$, where M is a set, $1 \in M$, and $\cdot : M \times M \to M$ is an operation such that for all $a, b, c \in M$

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c \qquad\qquad 1 \cdot a = a \cdot 1 = 1$$

We usually call this operation *multiplication*, and omit the $\cdot$ when writing it. We also denote monoids with just their set, M. A monoid is called *Abelian* if its multiplication is commutative.

Example 2.1.1. The natural numbers $\mathbb{N}$ under addition or multiplication are a monoid. Another good example is $\mathbb{R}^{n\times n}$ under multiplication, which is not an Abelian monoid.

Definition 2.1.2. Let M be a monoid. For any $a \in M$, we call $b \in M$ the right-inverse of a if $ab = 1$, and the left-inverse if $ba = 1$. b is called the inverse of a if it is both a left and right inverse. A monoid where every element has an inverse is called a group.

Example 2.1.2. $\mathbb{R}^{n\times n}$ under multiplication is not a group, due to the lack of inverses. The set of all invertible matrices in $\mathbb{R}^{n\times n}$ is a group.

Theorem 2.1.3. *Let $a \in M$ be an invertible (has an inverse) element of a monoid. Then its inverse is unique.*

Proof. Suppose $b, c \in M$ are inverses of a. Then $ab = ac \Rightarrow b(ab) = b(ac) \Rightarrow (ba)b = (ba)c \Rightarrow 1b = 1c \Rightarrow b = c$. □

As a result of Theorem 2.1.3, we denote the unique inverse of an element $a \in M$ by a^{-1}.

Definition 2.1.4. A sub-monoid (sub-group) of a monoid M is a subset of M which is itself a monoid (group).

Theorem 2.1.5. *Let $\{U_\alpha\}_{\alpha\in I}$ be a collection of sub-monoids of a monoid M. Then $U = \bigcap_{\alpha\in I} U_\alpha$ is a sub-monoid of M.*

Note. This is also equivalent to our notion of the algebra generated by a set from section 6.2.

Proof. Since $1 \in U_\alpha$ for every $\alpha \in I$, $1 \in U$. Associativity of multiplication is inherited from the monoid M, so it suffices to check that U is closed under multiplication. Pick any $a, b \in U$. Then $a, b \in U_\alpha$ for all $\alpha \in I$, so $ab \in U_\alpha$ for all $\alpha \in I$. Thus, $ab \in U$, as was to be shown. □

Note. An identical result holds for groups.

Definition 2.1.6. Let $S \subseteq M$ be a subset of a monoid M. The sub-monoid generated by S, denoted by $\langle S \rangle$, is the intersection of all sub-monoids of M containing S.

Note. By the result of Theorem 2.1.5, $\langle S \rangle$ is in fact a sub-monoid of M.

We next state a result which provides a much more practical way of expressing the module generated by a set.

Theorem 2.1.7. *$\langle S \rangle$ is the set of all elements of M that may be written as a product of 1 and the elements of S.*

Proof. Let X be the collection of products of elements of S and 1. Since $\langle S \rangle$ is a sub-monoid containing S and 1, any product of those elements is in $\langle S \rangle$, so $X \subseteq \langle S \rangle$. But X is closed under multiplication, and hence a sub-monoid of M containing S. Therefore, $\langle S \rangle \subseteq X$, as was to be shown. □

Note. We can make an identical defintion for groups. Theorem 2.1.7 still holds if you include the inverse of every element in S in the products.

Definition 2.1.8. Let $a \in M$ be an element of a monoid. The order of a, denoted $o(a)$, is the smallest $n \in \mathbb{N}$ such that $a^n = 1$. If no such natural number exists, we write that $o(a) = \infty$.

Theorem 2.1.9. *If an element $a \in M$ has finite order, it is invertible.*

Proof. Suppose that $o(a) = n$. Then $a^n = 1 = a^{n-1}a = aa^{n-1}$, so $a^{n-1} = a^{-1}$. □

Theorem 2.1.10. *Every element of a finite group G has finite order.*

Proof. Suppose $a \in G$ has infinite order. Pick any $n \geq m \in \mathbb{N}$. If $a^n = a^m$, then $a^{n-m} = 1$. Then a has finite order if $n \neq m$, so $n = m$. It follows that a^k is distinct for each $k \in \mathbb{N}$. But this would imply that G is infinite, a contradiction. □

Definition 2.1.11. Let M, N be monoids. A homomorphism $\varphi : M \to N$ is a map such that for any $a, b \in M$, $\varphi(ab) = \varphi(a)\varphi(b)$.

Note. Since $\varphi(a) = \varphi(1a) = \varphi(1)\varphi(a)$ and $\varphi(a) = \varphi(a1) = \varphi(a)\varphi(1)$ for any homomorphism φ and monoid element a, $\varphi(1) = 1$ for any homomorphism.

An injective homomorphism is called a *monomorphism*, and a surjective one an *epimorphism*. A bijective homomorphism is called an *isomorphism*. If there exists an isomorphism between two monoids M, N, we call them *isomorphic* and write that $M \cong N$. Note that $\cong$ is an equivalence relation.

2.2 Group of Transformations

This section is partially based on (and uses some proofs from) a similar section in [Jac09]. For the next few definitions, let X be an arbitrary set.

Definition 2.2.1. The monoid $M(X)$ of all transformations on X is the set of all set functions $f : X \to X$, with multiplication given by function composition. The group $\mathrm{Sym}(X)$ of all invertible elements of $M(X)$ is the symmetric group of X.

It is not too difficult to check that these are indeed a monoid and group respectively. Any sub-monoid $M(X)$ is called a monoid of transformations, and any sub-group of $\mathrm{Sym}(X)$ a group of transformations.

Definition 2.2.2. Suppose that X is a finite set with $|X| = n$. Then we call $\mathrm{Sym}(X) = S_n$ the permutation group on n elements. A sub-group of S_n is called a permutation group on n elements.

We generally represent permutations $\sigma \in S_n$ as the product of disjoint cycles. That is, denoting $X = \{1, 2, \ldots, n\}$ we first look at the sequence

$$1, \sigma(1), \sigma^2(1), \ldots$$

Since $|X| < \infty$, this sequence is finite. Since σ is invertible, the image of the final element of the sequence is 1. We call this sequence a disjoint cycle, and write it $(1\sigma(1)\sigma^2(1)\cdots)$. This represents a function on X that takes 1 to $\sigma(1)$, $\sigma(1)$ to $\sigma^2(1)$ and so on, and does nothing on any element not in the cycle. Repeating this process with an element not in any previous cycle, we get the following result.

Proposition 2.2.3. *Any permutation can be written as the product of disjoint cycles, and this product is unique up to the order of the cycles. Furthermore, disjoint cycles commute.*

Note. The presentation of any given cycle is not unique, for example $(123) = (312) = (231)$. In general, doing a cyclic permutation of the elements of a cycle does not change the cycle that it represents.

Now, we present two extremely important results on symmetric groups.

Theorem 2.2.4. *For any $n \in \mathbb{N}$, $|S_n| = n!$.*

Proof. We do this constructively. Pick any $\sigma \in S_n$. We have n valid choices for $\sigma(1)$. Then since σ is bijective, we have $n - 1$ for $\sigma(2)$, $n - 2$ for $\sigma(3)$, and so on until we have only one choice for $\sigma(n)$. Thus, $|S_n| = n!$, as was to be shown. □

Theorem 2.2.5 (Cayley's Theorem)**.** *Let G be a finite group. Then there exists a finite set X and group of transformations H on X such that $G \cong H$.*

Proof. Set $X = G$, and for any $g \in G$ consider the transfomation $L_g : X \to X$ given by

$$L_g(x) = gx$$

We'll prove that $H = \{L_g \mid g \in G\}$ is a group, and that the map $\varphi : g \mapsto L_g$ is an isomorphism. Let $L_a, L_b, L_c \in H$. Then we get

$$\begin{aligned}(L_aL_1)(x) = L_a(L_1(x)) = L_a(1x) = L_a(x) = ax\\ (L_1L_a)(x) = L_1(L_a(x)) = L_1(ax) = 1(ax) = ax\\ (L_aL_b)(x) = L_a(L_b(x)) = L_a(bx) = a(bx) = (ab)x = L_{ab}(x)\\ ((L_aL_b)L_c)(x) = (L_aL_b)(cx) = (ab)(cx) = a(bcx) = (L_a(L_bL_c))(x)\\ (L_aL_{a^{-1}})(x) = L_a(L_{a^{-1}}(x)) = a(a^{-1}x) = x\\ (L_{a^{-1}}L_a)(x) = L_{a^{-1}}(L_a(x)) = L_{a^{-1}}(ax) = a^{-1}(ax) = x\end{aligned}$$

The first two lines prove that L_1 is our unit 1, the third closure under multiplication, the fourth associativity, and the last two that $L_{a^{-1}}$ is the inverse of L_a, so H is a group. Suppose that $a, b \in G$ satisfy $\varphi(a) = \varphi(b)$. Then in particular, $b = b1 = \varphi(b)(1) = \varphi(a)(1) = a1 = a$, so φ is injective. φ is clearly surjective, and is hence bijective. It remains to show that it is a homomorphism. Indeed, we get for any $x \in G$

$$\varphi(ab)(x) = (ab)(x) = a(bx) = \varphi(a)(\varphi(b)(x)) = (\varphi(a)\varphi(b))(x)$$

so φ is a homomorphism, completing the proof. □

Note. Cayley's theorem allows us to study any finite group (in theory) by studying subgroups of S_n, which simplifies our life considerably. The only trick is that you often need to understand the structure of the group to find its corresponding subgroup of S_n, which has an unfortunate circular quality.

Note. Cayley's theorem extends to infinite monoids/groups, which are isomorphic to monoids/groups of transformation. The proof is essentially identical.

Finally, we'll develop the notion of the *sign* of a permutation. Before starting on this journey, we'll need the following lemma.

Lemma 2.2.6. *Let $a, b, c_1, \ldots, c_m, d_1, \ldots d_k$ be distinct elements in a finite set X. Then the following two equations hold.*

1. $$(ab)(ac_1 \cdots c_m bd_1 \cdots d_k) = (ac_1 \cdots c_m)(bd_1 \cdots d_k)$$

2. $$(ab)(ac_1 \cdots c_m)(bd_1 \cdots d_k) = (ac_1 \cdots c_m bd_1 \cdots d_k)$$

Proof. The first of these is obtained by tracing through the result of applying both sides to $a, b, c_1, \ldots, c_m, d_1, \ldots d_k$. We also note that $(ab)^2 = 1$. Then applying (ab) to both sides of the second equality, we see that it is in fact equivalent to the first. □

Note. Result (1) in the proceeding lemma implies that any disjoint cycle can be split into two smaller disjoint cycles. The above results also still hold if $m = 0$ or $k = 0$.

Using this, we can find an alternative way of representing permutations.

Theorem 2.2.7. *Any permutation can be written as the product of transpositions (cycles of the form (ab), where $a \neq b$).*

Proof. Since every permutation can be written as the product of disjoint cycles, it suffices by Lemma 2.2.6 to prove this for cycles of the form (abc), where a, b, c are all distinct. Indeed, we note that $(abc) = (ab)(bc)$, as required. □

This transposition decomposition is closely related to the sign of a permutation, which we now (finally) define.

Definition 2.2.8. Let $\sigma = C_1 C_2 \cdots C_m$ be the disjoint cycle representation of a permutation σ, where each C_i is a cycle of length d_i. Then the sign of σ is defined as

$$\operatorname{sgn}(\sigma) = (-1)^{\sum_{i=1}^{r}(d_i - 1)}$$

Note. Since the disjoint cycle decomposition of a permutation is unique up to the order of cycles, sgn is in fact well-defined.

Proposition 2.2.9. *Let σ be a permutation and τ a transposition. Then* $\operatorname{sgn}(\tau\sigma) = -\operatorname{sgn}(\sigma)$.

Proof. Let $\tau = (ab)$. There are then two cases to consider.
Case 1 : Suppose that a, b are in two different cycles in σ. Without loss of generality (since we can just cyclically permute the elements of a cycle, and we can permute disjoint cycles), these cycles are $C_1 = (ac_1 \cdots c_m)$ and $C_2 = (bd_1 \cdots d_k)$. Then writing $\sigma = C_1 C_2 \cdots C_n$ as a product of disjoint cycles with lengths l_k, we get by Lemma 2.2.6 the following

$$\tau\sigma = (ab)(ac_1 \cdots c_m)(bd_1 \cdots d_k)C_3 \cdots C_n = (ac_1 \cdots c_m bd_1 \cdots d_k)C_3 \cdots C_n$$

This new cycle is disjoint from the other C_i, and has length $l_1 + l_2$. Thus, we get

$$\operatorname{sgn}(\tau\sigma) = (-1)^{(l_1 + l_2 - 1) + \sum_{i=3}^{n}(l_i - 1)} = (-1)^{1 + \sum_{i=1}^{n}(l_i - 1)} = -\operatorname{sgn}(\sigma)$$

as required.
Case 2 : Suppose that a, b are in the same cycle in σ. Without loss of generality, we write this cycle as $C_1 = (ac_1 \cdots c_m bd_1 \cdots d_k)$. Then writing $\sigma = C_1 \cdots C_n$ as a product of disjoint cycles with lengths l_k, we get by Lemma 2.2.6 the following

$$\tau\sigma = (ab)(ac_1 \cdots c_m bd_1 \cdots d_k)C_2 \cdots C_n = (ac_1 \cdots c_m)(bd_1 \cdots d_k)C_2 \cdots C_n$$

These two new cycles are disjoint from the other C_i, and have lengths $l_{11} + l_{12} = l_1$. Thus, we get

$$\operatorname{sgn}(\tau\sigma) = (-1)^{(l_{11} - 1) + (l_{12} - 1) + \sum_{i=3}^{n}(l_i - 1)} = (-1)^{\sum_{i=1}^{n}(l_i - 1) - 1} = -\operatorname{sgn}(\sigma)$$

as required. □

This proposition turns out to be quite useful in proving the following theorem.

Theorem 2.2.10. *For any permutations σ_1, σ_2,* $\operatorname{sgn}(\sigma_1\sigma_2) = \operatorname{sgn}(\sigma_1)\operatorname{sgn}(\sigma_2)$.

Proof. By Theorem 2.2.7, we may write $\sigma_1 = \tau_1 \cdots \tau_n$ as a product of transposition. Then by repeated application of Proposition 2.2.9 we get

$$\operatorname{sgn}(\sigma_1\sigma_2) = (-1)^n \sigma \operatorname{sgn}(\sigma_2) = \operatorname{sgn}(\sigma_1)\operatorname{sgn}(\sigma_2)$$

as required. □

This finally brings us to the full relation between transpositions and the sign of a permutation, which is an immediate consequence of Theorem 2.2.10.

Corollary 2.2.10.1. *Suppose a permutation σ can be written as the product of k transpositions. Then $\operatorname{sgn}(\sigma) = (-1)^k$. Furthermore, any representation of σ as a product of transpositions must have an even amount of transpositions if $\operatorname{sgn}(\sigma) = 1$, and an odd amount otherwise.*

2.3 Cosets and Quotient Groups

Definition 2.3.1. Let G be a group, and $H \subseteq G$ a subgroup. We define two equivalence relations $\sim_L, \sim_R$ by, for any $x, y \in G$

$$x \sim_L y \iff \exists h \in H \mid hx = y, x \sim_R y \iff \exists h \in H \mid xh = y$$

We call the equivalence classes in $G/\sim_L, G/\sim_R$ left and right cosets respectively. They are given more explicitly by, for any $x \in G$

$$xH = \{xh \mid h \in H\}, Hx = \{hx \mid h \in H\}$$

Note. While not proven in the definition, it is not too hard to prove that these are indeed equivalence relations.

What we're really after here is some unified notion of the quotient group, so we want to find out what left and right cosets have in common.

Lemma 2.3.2. *Let G be a group, and $H \subseteq G$ a subgroup. For any $x \in G$, $|xH| = |Hx| = |H|$.*

Proof. Consider the function $f : H \to xH$ given by $f(h) = xh$. This is surjective, and injective since $f(h_1) = f(h_2) \Rightarrow xh_1 = xh_2 \Rightarrow h_1 = h_2$ (by cancelling the x on both sides). Thus, it is bijective. The proof for Hx is identical. □

This gives an immediate corollary.

Corollary 2.3.2.1. *If G is finite, then for any subgroup $H \subseteq G$ there are the same number of left and right cosets.*

We use this to build a new definition.

Definition 2.3.3. Let G be a finite group and $H \subseteq G$ a subgroup. We denote the number of left/right cosets (xH or Hx) in G by $[G : H]$, and call it the index of H in G.

Which brings us to our first big result of the section.

Theorem 2.3.4 (Lagrange's Theorem)**.** *Let G be a finite group and $H \subseteq G$ a subgroup. Then $|G| = [G : H]|H|$.*

Proof. By lemma 2.3.2, each coset xH (we choose left cosets here for convienience) in G has size $|H|$. The result is then immediate from the definition of $[G : H]$. □

This has a couple of immediate corollaries.

Corollary 2.3.4.1. *Let $g \in G$ be any element of a finite group. Then $o(g) \mid |G|$.*

Proof. It suffices to note that $o(g) = |\langle g \rangle|$. □

Corollary 2.3.4.2. *Let $g \in G$ be any element of a finite group. Then $g^{|G|} = 1$.*

Proof. Since $o(g) \mid |G|$, $\exists k \in \mathbb{N}$ such that $o(g)k = |G|$. Thus

$$g^{|G|} = (g^{o(g)})^k = 1^k = 1$$

□

We now move on towards quotient groups, for which we really want left and right cosets to be identical.

Definition 2.3.5. A subgroup $H \subseteq G$ is called normal, denoted $H \trianglelefteq G$, if for all $x \in G$, $xH = xH$.

When $H \trianglelefteq G$, we often just refer to the coset of an element, and don't bother specifying left or right.

Theorem 2.3.6. *$H \trianglelefteq G$ if and only if for all $h \in H$ and $g \in G$, $ghg^{-1} \in H$.*

Proof. Suppose $H \trianglelefteq G$, and pick any $h \in H$. Then for any $g \in G$, $gh \in Hg$. Thus, $\exists h' \in H$ such that $gh = h'g \Rightarrow ghg^{-1} \in H$, as was to be shown. Now, suppose that $ghg^{-1} \in H$ for all $g \in G, h \in H$. Let $h' = ghg^{-1}$. Then $h'g = gh$, so $gh \in Hg$. Since this was for arbitrary g, h, this implies that $gH = Hg$, as was to be shown. □

Theorem 2.3.7. *Suppose $H \trianglelefteq G$. Then the set of cosets of G relative to H, with multiplication defined by $(xH)(yH) = (xy)H$, is a group.*

Proof. All the properties of group multiplication are inherited from G if this multiplication is well-defined, so we just need to check this. Suppose $xH = yH$ and $aH = bH$. Then since H is normal, $\exists h, h' \in H$ such that $hx = y, ah' = b$ (this is using the same technique as in the proof of Theorem 2.3.6, just taking one of the values in H to be 1). Since multiplying by elements of H doesn't change the coset, it follows that

$$(xH)(aH) = (xa)H = (xah')H = (xb)H = H(xb) = H(hxb) = H(yb) = (yb)H$$

completing the proof. □

Definition 2.3.8. The group from Theorem 2.3.7 is denoted G/H, and called the quotient group of G by H, or G mod H.

Note. When G/H is a quotient group, the canonical quotient map becomes a homomorphism.

It's worth taking a moment here to reflect on how this connects to section 6.1. The idea here is that equivalence relations on groups are congruences if and only if the equivalence classes are cosets of some normal subgroup of the group.

2.4 Homomorphism Theorems

Here, we prove the group analogs of the homomorphism theorems in section 6.1. These would all (technically) follow immediately from the results there, but for the sake of those who aren't as familiar with the material I will present direct proofs of the theorems. Again, this is a reworking of a similar section in [Jac09].

Definition 2.4.1. Let $\varphi : G_1 \to G_2$ be a group homomorphism. The kernel of φ is given by

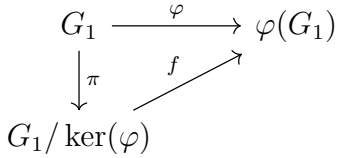
$$\ker(\varphi) = \{g \in G_1 \mid \varphi(g) = 1\}$$

Lemma 2.4.2. $\ker(\varphi) \trianglelefteq G_1$.

Proof. We know that $\varphi(1) = 1$, so this condition is satisfied. It remains to check closure under multiplication and the existence of multiplicative inverses. For the former, we note that if $x, y \in \ker(\varphi)$, then $\varphi(xy) = \varphi(x)\varphi(y) = 1$, so $xy \in \ker(\varphi)$. For the latter, we note that for any $x \in G_1$, $\varphi(xx^{-1}) = 1 = \varphi(x)\varphi(x^{-1})$, so $\varphi(x)^{-1} = \varphi(x^{-1})$. Thus, if $x \in \ker(\varphi)$, $\varphi(x^{-1}) = \varphi(x)^{-1} = 1^{-1} = 1$, so $x^{-1} \in \ker(\varphi)$. □

We denote the set of all homomorphisms between two groups G_1, G_2 by $\mathrm{Hom}(G_1, G_2)$. If the homomorphism is between a group and itself, we drop the second group in that notation. We denote all the isomorphisms by $\mathrm{Isom}(G_1, G_2)$.

Theorem 2.4.3 (First Fundamental Theorem of Homomorphisms)**.** *Let G_1, G_2 be groups and pick any $\varphi \in \mathrm{Hom}(G_1, G_2)$. Let $\pi \in \mathrm{Hom}(G_1, G_1/\ker(\varphi))$ be the quotient map. Then there exists an isomorphism $f \in \mathrm{Isom}(G_1/\ker(\varphi), \varphi(G_1))$ which makes the following diagram commute.*



Proof. We proceed by directly constructing f. Pick any $g\ker(\varphi) \in G_1/\ker(\varphi)$. We define that $f(g\ker(\varphi)) = \varphi(g)$. This is clearly a homomorphism if it is well-defined, so we prove that it is in fact well-defined. Suppose that $x, y \in G$ are elements such that $x\ker(\varphi) = y\ker(\varphi)$. Then $\exists h \in \ker(\varphi)$ such that $xh = y$, so $\varphi(y) = \varphi(xh) = \varphi(x)\varphi(h) = \varphi(x)$, as required. f is by definition surjective onto $\varphi(G_1)$, so it remains only to prove that it is injective. Suppose that $f(x\ker(\varphi)) = f(y\ker(\varphi))$. Then $\varphi(x) = \varphi(y) \Rightarrow \varphi(x)\varphi(y)^{-1} = 1$, so $\varphi(xy^{-1}) = 1$. Then $xy^{-1} \in \ker(\varphi)$, so $x\ker(\varphi) = y\ker(\varphi)$, as was to be shown. □

Corollary 2.4.3.1. *Any $\varphi \in \mathrm{Hom}(G_1, G_2)$ is injective if and only if $\ker(\varphi) = \{1\}$.*

Proof. Suppose φ is injective. Then by definition, $G_1 \cong \varphi(G_1)$. But by Theorem 2.4.3, $\varphi(G_1) \cong G_1/\ker(\varphi)$, so $G_1 \cong G_1/\ker(\varphi)$. Thus, the projection map $\pi : G_1 \to G_1/\ker(\varphi)$ is injective, so $\pi^{-1}(\ker(\varphi)) = \ker(\varphi) = \{1\}$, as required. Now, suppose that $\ker(\varphi) = \{1\}$. Then $x\ker(\varphi) = y\ker(\varphi)$ if and only if $x = y$, so π is injective. By Theorem 2.4.3, the following diagram commutes

$$\begin{array}{ccc} G_1 & \xrightarrow{\varphi} & \varphi(G_1) \\ \downarrow{\scriptstyle \pi} & \nearrow{\scriptstyle f} & \\ G_1/\ker(\varphi) & & \end{array}$$

where f is an isomorphism, so it follows that φ is injective. □

Lemma 2.4.4. *The image or pre-image of a subgroup under a homomorphism is a subgroup.*

Proof. Let G, H be groups and $\varphi : G \to H$ a homomorphism. Let $K \subseteq G$ be a subgroup. Then $1 \in K$, so $\varphi(1) = 1 \in \varphi(K)$. Pick any $a, b \in \varphi(K)$. Then $\exists x, y \in K$ such that $\varphi(x) = a, \varphi(y) = b$, so $\varphi(xy) = ab \in \varphi(K)$, making it a subgroup of H. Suppose $K \subseteq H$ is a subgroup. Then $1 \in K$, so $1 \in \varphi^{-1}(K)$ as required. Suppose $x, y \in \varphi^{-1}(K)$. Then $\varphi(xy) = \varphi(x)\varphi(y) \in K$, so $xy \in \varphi^{-1}(K)$ making it a group. □

Theorem 2.4.5 (Second Fundamental Theorem of Homomorphisms)**.** *Let G_1, G_2 be groups and $\varphi : G_1 \to G_2$ a surjective homomorphism. Then*

1. *There exists a bijection between all subgroups of G_1 containing $\ker(\varphi)$ and all subgroups of G_2.*
2. *If $H \supseteq \ker(\varphi)$ is a subgroup of G_1, $H \trianglelefteq G_1$ if and only if $\varphi(H) \trianglelefteq G_2$.*
3. *If $\ker(\varphi) \subseteq H \trianglelefteq G_1$, then $G_1/H \cong G_2/\varphi(H)$.*

Proof. Let $\pi : G_1 \to G_1/\ker(\varphi)$ be the quotient map. Then by lemma 2.4.4, π maps the subgroups of G_1 onto all subgroups of $G_1/\ker(\varphi)$. By Theorem 2.4.3, $G_1/\ker(\varphi) \cong G_2$, so by lemma 2.4.4 the isomorphism $f : G_1/\ker(\varphi) \to G_2$ is a bijective map between subgroups of $G_1/\ker(\varphi)$ and subgroups of G_2. Since the pre-image of any subgroup of $G_1/\ker(\varphi)$ under π contains $\ker(\varphi)$, $f \circ \pi$ is a surjection between subgroups of G_1 containing $\ker(\varphi)$ and subgroups of G_2. Finally, suppose that H_1, H_2 are subgroups of G_1 containing the kernel and $\pi(H_1) = \pi(H_2)$. Then $H_1 = \{h\ker\varphi\}_{h\in H_1} = \{h\ker\varphi\}_{h\in H_2} = H_2$, so $f \circ \pi = \varphi$ is injective between the sets of subgroups, making it a bijection.
Let $H \supseteq G_1$ contain the kernel. Suppose it is normal. Pick any $h \in \varphi(H), g \in G_2$. Then $\exists g' \in G_1, h' \in H$ such that $\varphi(g') = g, \varphi(h') = h$. Thus, $ghg^{-1} = \varphi(g')\varphi(h')\varphi(g'^{-1}) = \varphi(g'h'g'^{-1})$. Since H is normal, $g'h'g'^{-1} \in H$, so $ghg^{-1} \in \varphi(H)$, making it normal. Suppose, conversely, that $\varphi(H)$ is normal. Pick any $g \in G_1, h \in H$. Note, by part (1) of this theorem, that $\varphi^{-1}(\varphi(H)) = H$. Thus, $\varphi(ghg^{-1}) = \varphi(g)\varphi(h)\varphi(g)^{-1} \in \varphi(H)$ implies that $ghg^{-1} \in H$, as required.

Finally, suppose that $\ker(\varphi) \subseteq H \trianglelefteq G_1$. By part (2), we know that $\varphi(H) \trianglelefteq G_2$. Let $\pi' : G_2 \to G_2/\varphi(H)$ be the quotient map, and let $\psi = \pi' \circ \varphi$. Then by Theorem 2.4.5, $G_1/\ker(\psi) \cong G_2/\varphi(H)$. But $\ker(\psi) = \varphi^{-1}(\pi'^{-1}(1)) = \varphi^{-1}(\varphi(H)) = H$, so $G_1/H \cong G_2/\varphi(H)$. □

Definition 2.4.6. Let $H, K \subseteq G$ be subgroups. We define the product of the subgroups as

$$HK = \{hk \mid h \in H, k \in K\}$$

Lemma 2.4.7. *If $K \trianglelefteq G$, then HK is a subgroup of G.*

Proof. Since $1 \in H, K$, $1 \in HK$. Pick any $hk, h'k' \in HK$. Then

$$(hk)(h'k') = (hh')(h'^{-1}kh')k' = h''k''$$

for some $h'' \in H, k'' \in K$, as required. We just need to check inverses.

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

for some $k' \in K$, as required. □

Theorem 2.4.8 (Third Fundamental Theorem of Homomorphisms)**.** *Let G_1, G_2 be groups, $\varphi : G_1 \to G_2$ a surjective homomorphism, $K = \ker(\varphi)$, and $H \subseteq G_1$ any subgroup. Then*

$$\varphi(H) \cong \frac{HK}{K} \cong \frac{H}{H \cap K}$$

Proof. Let $\psi = \varphi_{\restriction HK}, \vartheta = \varphi_{\restriction H}$. Then since $K \subseteq HK$, $\psi^{-1}(1) = K$ and $\vartheta^{-1}(1) = K \cap H$. Furthermore, we can note that since $\psi(hk) = \phi(hk) = \phi(h)$, $\psi(HK) = \varphi(H)$. Thus, by Theorem 2.4.3 we have the following two commutative diagrams.

$$\begin{array}{ccc} HK & \xrightarrow{\psi} & \varphi(H) \\ \downarrow & \nearrow_{\cong} & \\ HK/K & & \end{array} \qquad \begin{array}{ccc} H & \xrightarrow{\vartheta} & \varphi(H) \\ \downarrow & \nearrow_{\cong} & \\ H/(K \cap H) & & \end{array}$$

as required. □

Note. Many books will put these theorems in a different order, give them different names, or add/remove conclusions from each. I do not claim that these are **the** authoritative correct fundamental theorems of homomorphisms.

2.5 Cyclic Groups

This section is essentially that of the same name presented in the first chapter of [Jac09], and aims to introduce properties of the simplest kind of groups; cyclic groups. I've done my best to rework it in a way that hopefully increases clarity.

Definition 2.5.1. A cyclic group is a group generated by a single element.

We refer to an element that generates a cyclic group as a *generator* of that group. In general, a cyclic group will have many possible generators.

Lemma 2.5.2. *If G is abelian, every subgroup of G is normal.*

Proof. If $K \subseteq G$, $k \in K, g \in G$, then $gkg^{-1} = gg^{-1}k = k \in K$. □

Theorem 2.5.3. *Suppose G is cyclic. Then if G is infinite, $G \cong \mathbb{Z}$ (the additive group of integers) and otherwise $G \cong \mathbb{Z}/|G|\mathbb{Z}$ (the additive group of integers modulo $|G|$).*

Proof. Let $G = \langle g \rangle$. If G is infinite, we can define a homomorphism $\varphi : G \to \mathbb{Z}$ by $g \mapsto 1$. Since $\varphi(g^n) = n\varphi(g) = n$, φ is injective and surjective, so $G \cong \mathbb{Z}$. Suppose $|G| = n$, that is $o(g) = n$. Then we can define a homomorphism $\varphi : G \to \mathbb{Z}/n\mathbb{Z}$ by $\varphi(g) = [1]$. This is certainly surjective. $\varphi(g^r) = \varphi(g^m) \Rightarrow [r] = [m] \Rightarrow r = kn + m$ for some $k \in \mathbb{Z}$. But then $g^r = g^{kn+m} = g^m$, so φ is injective. Finally, we note that since $g^n = 1$, $g^m = g^{m \bmod n}$ for any $m \in \mathbb{Z}$, so this map is well-defined. Thus, $G \cong \mathbb{Z}/n\mathbb{Z}$. □

Combining this with the transitivity of being isomorphic, we get an immediate corollary.

Corollary 2.5.3.1. *Any two cyclic groups of the same order are isomorphic.*

Theorem 2.5.4. *Any subgroup of a cyclic group is cyclic. If the cyclic group is infinite, the set of all non-trivial subgroups is in bijection with $\mathbb{N}$. If $G = \langle g \rangle$, where $o(g) = n$, then there is one and only subgroup of order q for every $q \mid n$.*

Proof. Let $G = \langle g \rangle$. For any subgroup $H \subseteq G$, there exists some smallest $n > 0$ such that $g^n \in H$ (note we can assume $n > 0$ since if $n < 0$, we simply take its inverse). Any element of the form g^m, where $n \mid m$, can be generated by this element. Suppose $\exists g^m \in H$ such that $n \nmid m$. Write $m = kn + r$, where $k \in \mathbb{Z}, 0 < |r| < n$. Then $g^{m-kn} = g^r \in H$. But $r < n$, contradicting the minimality of n. Hence, $H = \langle g^n \rangle$, as was to be shown.
Now, suppose that $o(g) = \infty$. Then $\langle g^n \rangle \subseteq G$ is a subgroup for each $n \in \mathbb{N}$. Furthermore, all non-trivial subgroups are of this form, as if the generator is g^n for $n < 0$ we simply take its inverse, and $g^0 = 1$. It suffices then to show that each of these is unique. Suppose that $\langle g^n \rangle = \langle g^m \rangle$ for some $m, n \in \mathbb{N}$. Then there exists some k such that $g^{mk} = g^n \Rightarrow g^{mk-n} = 1$. There also exists some $r \in \mathbb{N}$ such that $g^{rn} = g^m \Rightarrow g^{rn-m} = 1$. Since g has infinite order, $mk - n, rn - m = 0 \Rightarrow n \mid m, m \mid n$, so $m = n$ as required.
Suppose $o(g) = n$, and pick any subgroup $H \subseteq G$. Let $m \in \mathbb{N}$ be the minimal number such that $g^m \in H$. From the above, we know that $H = \langle g^m \rangle$. Pick any $q \mid n$. For existence, it suffices to note that $o(g^{n/q}) = q$. For uniqueness, let $m \in \mathbb{N}$ be the minimal number such that $o(g^m) = q$. Suppose $o(g^k) = q$. Then $\exists a, b \in \mathbb{N}$ such that $qm = an, qk = bn$. Then $m = \frac{an}{q}, k = \frac{bm}{q}$. Since m is minimal, it follows that $a = 1$ (as $a = 1$ certainly works), so $g^k \in \langle g^m \rangle$, as required. □

Corollary 2.5.4.1. *Let G be a group, and $a \in G$ an element such that $o(a) < \infty$. Then $\langle a \rangle = \{g \in G \mid o(g) \mid o(a)\}$.*

Proof. Suppose $g \in \langle a \rangle$. Then $g = a^n$ for some $n \in \mathbb{N}$. Thus, $o(g) \mid o(a)$, as required. Now, suppose that $o(g) \mid o(a)$. By Theorem 2.5.4, there is exactly one group of order $o(g)$, and g must generate this group. But $o(a^{o(a)/o(g)}) = o(g)$, so it follows that this group lies in $\langle a \rangle$, and in particular $g \in \langle a \rangle$. □

Lemma 2.5.5. *Let a, b be elements of an abelian group of finite orders $o(a) = n, o(b) = m$ such that $(n, m) = 1$. Then $\langle a \rangle \cap \langle b \rangle = \langle 1 \rangle$, $\langle a, b \rangle = \langle ab \rangle$, and $o(ab) = nm$.*

Proof. Suppose $x \in \langle a \rangle \cap \langle b \rangle$. Then $o(x) \mid n, m$, so $o(x) = 1 \Rightarrow x = 1$ as required. Since the group is abelian, $(ab)^k = a^k b^k$ for any $k \in \mathbb{N}$. In particular, if $k \in \mathbb{N}$ we get $a^k = b^{-k} \in \langle a \rangle \cap \langle b \rangle$, so $a^k = b^{-k} = 1 \Rightarrow n, m \mid k$. The smallest such k is nm, and $(ab)^{nm} = 1$, so $o(ab) = nm$. Finally, note that we can write every $x \in \langle a, b \rangle$ in the form $x = a^r b^q$, where $0 \leq r < n, 0 \leq q < m$. It follows that $|\langle a, b \rangle| \leq nm$. But $\langle ab \rangle \subseteq \langle a, b \rangle$, so $\langle a, b \rangle = \langle ab \rangle$. □

Lemma 2.5.6. *If G is a finite abelian group, then it contains an element whose order is divisible by the order of every element of G.*

Proof. Since G is finite, it suffices to show that we can do this with any two elements. Let $a, b \in G$, and take the prime decomposition of both orders

$$o(a) = n = p_1^{e_1} \cdots p_r^{e_r} \cdots p_l^{e_l} \qquad o(b) = m = p_1^{f_1} \cdots p_r^{f_r} \cdots p_l^{f_l}$$

We order things such that $e_i \geq f_i$ for $i \leq h$, and $f_i \geq e_i$ for $i > h$. We can see then that

$$[n, m] = p_1^{e_1} \cdots p_r^{e_r} p_{r+1}^{f_{r+1}} \cdots p_l^{f_l}$$

Let $q = p_1^{f_1} \cdots p_r^{f_r}, s = p_{r+1}^{e_{r+1}} \cdots p_l^{e_l}$. Then $[n, m] = \frac{n}{s}\frac{m}{q}$, and $(n/s, m/q) = 1$. $o(a^s) = n/s, o(b^q) = m/q$, so by lemma 2.5.5 $o(a^s b^q) = [n, m]$, as required. □

Theorem 2.5.7. *Let G be a finite abelian group. Then G is cyclic if and only if $|G|$ is the smallest positive integer n such that $a^n = 1$ for all $a \in G$.*

Proof. Suppose G is cyclic, with generator g. Then $o(g) = |G|$, so by Lagranges's theorem $a^{|G|} = 1$ for any $a \in G$. Since $o(g) = |G|$, this is the smallest such integer. Now, suppose that $|G|$ is the smallest positive integer n such that $a^n = 1$ for all $a \in G$. By lemma 2.5.6, $\exists g \in G$ whose order is divisible by the order of every element in G. Then $a^{o(g)} = 1$ for any $a \in G$, and $o(g) \leq |G|$, so $o(g) = |G|$. Thus, $G = \langle g \rangle$, completing the proof. □

2.6 Group Actions

This section is based on similar sections in [Jac09] and [Lan05]. In it, we introduce one of the most important applications of groups in mathematics, *group actions*.

Definition 2.6.1. An action of a group G on a set S is a homomorphism $f \in \mathrm{Hom}(G, \mathrm{Sym}(S))$.

If a group G acts on a set S, we call S a G-set. Before looking at examples, let's prove a result that makes the reasoning for calling this a group action more clear.

Theorem 2.6.2. *Let G be a group and S a set. S is a G-set if and only if there exists a map $\cdot : G \times S \to S$ satisfying the following axioms for any $x \in S$, $g, h \in G$*

1. $1 \cdot x = x$
2. $(gh) \cdot x = g \cdot (h \cdot x)$

Proof. Suppose S is a G-set. Then there exists a homomorphism $\varphi : \mathrm{Hom}(G, \mathrm{Sym}(S))$, in which case we simply define that $g \cdot x = \varphi(g)(x)$ (one can check that this has the desired properties). Now, suppose that the map $\cdot$ exists. Then $x \mapsto g \cdot x$ is a permutation of S, and $\varphi(g) = (x \mapsto g \cdot x)$ is the desired homomorphism. □

Note. The above version of a group action is also called a *left group action*, with *right group actions* being similar but with a map $\cdot S \times G \to S$. These are essentially identical to left group actions, so we won't be bothering with them here. They are occasionally a clearer notation though.

Example 2.6.1. Given any group G and subgroup $H \subseteq G$, G acts on G/H by $g \cdot (kH) = (gk)H$ (for any $g, k \in G$).

A group action is *effective* if the map $\varphi \in \mathrm{Hom}(G, \mathrm{Sym}(S))$ is injective. It is *faithful* if $g \cdot x$ for all $x \in S$ implies that $g = 1$, and *free* if $g \cdot x$ for some $x \in S$ implies that $g = 1$.

The rest of this section is admittedly a little haphazard, exploring the many different directions of inquiry we could take with group actions. Let's start by figuring out how to move between group actions.

Definition 2.6.3. Let X, Y be G-sets. A map $f : X \to Y$ is called a morphism of G-sets if for all $g \in G, x \in X$, $f(g \cdot x) = g \cdot f(x)$.

Two G-sets are said to be *isomorphic* or *equivalent* if there exists an invertible morphism between them, whose inverse is also a morphism of G-sets.

Let X be a G-set. The G-orbit of an element $x \in S$ is $Gx = \{g \cdot x \mid g \in G\}$. Its clear that S can be partitioned into disjoint G-orbits. A G-set S is called *transitive* (or the action of G on S called transitive) if S has exactly one G-set.

Definition 2.6.4. The stabilizer of an element $x \in X$ of a G-set X is defined as

$$\mathrm{Stab}_G(x) = \{g \in G \mid g \cdot x = x\}$$

Lemma 2.6.5. *For any $x \in X$, $Stab_G(x)$ is a subgroup of G.*

Proof. Suppose $a, b \in \mathrm{Stab}_G(x)$. Then $(ab) \cdot x = a \cdot (b \cdot x) = a \cdot x = x$, so $ab \in \mathrm{Stab}_G(x)$. $1 \in \mathrm{Stab}_G(x)$ is clear. Finally, we get

$$x = 1 \cdot x = (a^{-1}a) \cdot x = a^{-1} \cdot (a \cdot x) = a^{-1} \cdot x$$

so $a^{-1} \in \mathrm{Stab}_G(x)$. □

Using this, we can get a very strong result on transitive group actions.

Theorem 2.6.6. *Suppose X is a transitive G-set. For any $x \in S$, set $H = Stab_G(x)$. Then the action of G on X is equivalent to the left action of G on G/H, as given in example 2.6.1.*

Proof. Pick any $x \in X$. We'll define $f : X \to G/H$ by $f(g \cdot x) = gH$, for any $g \in G$ (note that this only works since X is transitive). First, we need to show that this is well-defined. Suppose $a, b \in G$ are such that $a \cdot x = b \cdot x$. Then $\exists g \in G$ such that $b = ag$, so $(ag) \cdot x = a \cdot (g \cdot x) \Rightarrow g \cdot x = 1 \cdot x = x$, and thus $g \in H$ and $aH = bH$. Next, we show that this is a morphism of G-sets. Pick any $x \in X, g \in G$. Then we get

$$f(g \cdot x) = gH = g \cdot (1H) = g \cdot f(1 \cdot x) = g \cdot f(x)$$

as required. Finally, we show that it is bijective. That f is surjective is clear from the definition. Suppose that f were not injective. Then $\exists a, b \in G$ such that $a \cdot x \neq b \cdot x$ but $aH = bH$. Thus, $\exists h \in H$ such that $b = ah$, so $b \cdot x = (ah) \cdot x = a \cdot (h \cdot x) = a \cdot x$, which is impossible. f is therefore bijective, as was to be shown. □

This leads immediately to a very useful corollary.

Corollary 2.6.6.1. *If X is a transitive G-set, then $|X| = [G : Stab_G(x)]$ for any $x \in X$.*

This is actually more powerful than it looks at first glance for group actions on finite sets. Suppose X is a finite G-set. Then for any $x \in X$, we can regard Gx as a finite G-set by restricting the action on X to Gx. It's clear in this case that G acts transitively on Gx. Furthermore, if X is finite, then it divides up into finitely many disjoint G-orbits. This, combined with corollary 2.6.6.1, gives us the following important result.

Theorem 2.6.7. *Let X be a finite G-set, and $\bigsqcup_{i=1}^{n} Gx_i = X$ be a decomposition of X into disjoint G-orbits. Then*

$$|X| = \sum_{i=1}^{n} [G : Stab_G(x_i)]$$

Proof. Since the Gx_i are disjoint, $|X| = \sum_{i=1}^{n} |Gx_i|$, and by corollary 2.6.6.1 we know that since G acts on Gx_i transitively, $|Gx_i| = [G : \mathrm{Stab}_G(x_i)]$. □

We now take a quick detour into the world of *primitive group actions*.

Definition 2.6.8. Let X be a G-set, and $\pi(X)$ a partition of X. We say that $\pi(X)$ is stabilized by the G-action if $g \cdot Y \in \pi(X)$ for all $g \in G, Y \in \pi(X)$.

Note. There are three partitions of a set X which will always have this property. The partition X, the partition $\{x\}_{x \in X}$ and the partition of X into G-orbits.

Definition 2.6.9. Let X be a G-set. We say that G acts primitively on X if the only two partitions of X stabilized by G are X and $\{x\}_{x \in X}$. Otherwise, we say it acts imprimitively.

Note. Since the partition of X into G-orbits is always stabilized by G, G acts primitively only if it acts transitively.

The next two results have proofs taken directly from [Jac09].

Lemma 2.6.10. *G acts imprimitively on a set X if and only if $\exists A \subsetneq X$ such that $|A| > 1$ and for any $g \in G$, either $g \cdot Y = Y$ or $(g \cdot Y) \cap Y = \emptyset$.*

Proof. Suppose there exists $Y \subsetneq X$ meeting the above conditions. Then for any $g_1, g_2 \in G$, $g_1 \cdot Y$ and $g_2 \cdot Y$ are either equal or disjoint. Let $Z = X \backslash (\bigcup_{x \in X} g \cdot Y)$. Then $g_1 \cdot B \cap g_2 \cdot Y = \emptyset$ for all $g_1, g_2 \in G$, so $g_1 \cdot B = B$ for all $g_1 \in G$. Thus, the set of all distinct subsets of the form $g \cdot Y$, along with B, forms a partition of X stabilized by G, making the action of G imprimitive. Now, suppose that G acts imprimitively on X. Then there exists a partition $\pi(X)$ which G stabilizes, which must contain some $Y \in \pi(X)$ such that $|Y| > 1, Y \subsetneq X$. Since this partition is stabilized, it follows that for any $g \in G$, either $g \cdot Y = Y$ or $(g \cdot Y) \cap Y = \emptyset$. □

Theorem 2.6.11. *Suppose G acts transitively on a set X, where $|X| > 1$. Then G acts primitively if and only if, for any $x \in X$, $Stab_G(x)$ is a maximal subgroup of G (i.e. there exists no group H such that $Stab_G(x) \subsetneq H \subsetneq G$).*

Proof. Suppose $\exists x \in X$ such that $\mathrm{Stab}_G(x)$ is not maximal, and let H be a subgroup such that $\mathrm{Stab}_G(x) \subsetneq H \subsetneq G$. Since G acts transitively, we get by Theorem 2.6.6 that the G-action on X is equivalent to the G-action on $G/\mathrm{Stab}_G(x)$, and thus G acts imprimitively on X if and only if it acts imprimitively on $G/\mathrm{Stab}_G(x)$. Let Y be the set of cosets of the form $h\mathrm{Stab}_G(x)$, where $h \in H$. Since $\mathrm{Stab}_G(x) \subsetneq H \subsetneq G$, $|Y| > 1$ and $Y \neq G/\mathrm{Stab}_G(x)$. It's also clear that $h \cdot Y = Y$ for any $h \in H$. If $g \notin H$, then since $gh \notin H$ for any $h \in H$ we have that $(g \cdot Y) \cap Y = \emptyset$. Thus, by lemma 2.6.10, G acts imprimitively on $G/\mathrm{Stab}_G(x)$ and hence on X.

Now, suppose that G acts transitively and imprimitively on X. Then by lemma 2.6.10, there exists a proper subset $Y \subsetneq X$ such that $|Y| > 1$, and for any $g \in G$, either $g \cdot Y = Y$ or $(g \cdot Y) \cap Y = \emptyset$. Let $H = \{h \in G \mid h \cdot Y = Y\}$. H is clearly a subgroup of G which contains $\mathrm{Stab}_G(x)$ for any $x \in Y$, since for any $g \in G$ we get $g \cdot x = x$ implies that $(g \cdot Y) \cap Y \neq \emptyset$, so $g \cdot Y = Y$. Since $Y \neq X$ and g acts transitively, $\exists g \in G$ such that $g \cdot Y \neq Y$. Thus, $g \notin H$, so $H \neq G$. Finally, let $x, y \in Y$ be distinct elements. Then $\exists g \in G$ such that $g \cdot x = y$. Thus, $g \in H, g \notin \mathrm{Stab}_G(x)$, so $H \neq \mathrm{Stab}_G(x)$, making $\mathrm{Stab}_G(x)$ not a maximal subgroup of G. □

Next, we look at possibly the most important kind of group action; *conjugation*.

Definition 2.6.12. Let G by any group. The action of G on itself by conjugation is given by, for any $g, x \in G$, $g \cdot x = gxg^{-1}$. The orbits of this action are called the *conjugacy classes* of G.

Example 2.6.2. The conjugacy classes of S_n are all the disjoint cycle decompositions of the same time, in the sense that two cycles are conjugate if and only if their decomposition has the same number of disjoint cycles of each size.

We also bring up now the centralizer $C(S)$ of a subset S of a group G, which is the set of all elements which commute with every element in G. This can be found to be a subgroup of G. $C(G) = C$ is called the centre of a group. For action by conjugation, it's clear that $\mathrm{Stab}_G(x) = C(x)$. Thus, we get by Theorem 2.6.7 that for finite groups

$$|G| = \sum_{i=1}^{n} [G : C(x_i)]$$

where $x_i \in G$ are representatives of the conjugacy classes of G. This is called the *class equation of a finite group*. We can also note that $C(x) = |G|$ and the conjugacy class of x is x for elements in C, so this is also often written as

$$|G| = |C| + \sum_{i=1}^{n} [G : C(y_i)]$$

where y_i are representatives of the conjugacy classes outside of C. We use this now to do a cute little proof which is often useful when working with finite groups.

Theorem 2.6.13. *Any finite group G of prime power order has a non-trivial centre.*

Proof. Let $|G| = p^n$, and let $y_i \in G$ be the representatives for the conjugacy classes outside of C. We know that $p \mid |G|$, and $p \mid |C(y_i)|$ for each y_i (as $|C(y_i)| \neq 1$ and $C(y_i)$ is a subgroup of G). Thus, by the class equation we must get $p \mid |C| \Rightarrow |C| \neq 1$. □

2.7 Free Groups*

This section will be much more informal than the rest, as a formal treatment of free groups requires delving into a level of category theory that is best left for a more advanced course in algebra. For a formal treatment, see [Lan05].

Definition 2.7.1. Let S be an arbitrary set. The free group on S, denoted F_S, is the group of finite strings of the elements x and x^{-1} for $x \in S$, along with the unit string 1. Multiplication is concatenation of strings, with the rule that $xx^{-1} = 1$.

Example 2.7.1. The free group on two elements, F_2 is the set of all strings of the letters a, b, a^{-1}, b^{-1}. In it, we'd have that $abb^{-1}a^{-1} = 1$, but this group is not commutative and $aba^{-1}b^{-1} \neq 1$ cannot be simplified.

Free groups are often used to specify arbitrary finitely generated groups in a simple manner. To show how this is done, we need the following definition.

Definition 2.7.2. Let G be a group, and $S \subseteq G$ a subset. The *normal subgroup generated by* S, denoted $\langle S \rangle_N$, is the intersection of all normal subsets of G containing S, or equivalently the smallest normal subgroup of G containing S.

There is unfortunately no simple way to write elements of a normal subgroup generated by a set, as there was for the regular subgroup generated by a set. However, there is a simple way to write elements of $G/\langle S \rangle_N$.

Definition 2.7.3. Let S be sets, and R a subset of F_S. The group with generators S and relations R is the free group S, with the addition rule enforced that any string in R is equal to 1. The group presentation of this group is $\langle S \mid R \rangle$.

Theorem 2.7.4. $\langle S \mid R \rangle \cong \frac{F_S}{\langle R \rangle_N}$.

Proof. We start with the natural surjective homomorphism $\varphi : F_S \to \langle S \mid R\rangle$. Then clearly $R \subseteq \ker(\varphi)$, so $\langle R\rangle_N \subseteq \ker(\varphi)$. Let $J = \ker(\varphi)\backslash\langle R\rangle$. Then $\varphi(r) = 1$ for any $r \in J$, so adding J to R wouldn't change $\langle S \mid R\rangle$. Thus, we must get $\ker(\varphi) = \langle R \cup J\rangle_N = \langle R\rangle_N$, so by the first fundamental theorem of homomorphisms $\langle S \mid R\rangle \cong \frac{F_S}{\langle R\rangle_N}$ as claimed. □

Example 2.7.2. The presentation of the group $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ is $\langle a, b \mid a^2, b^2, aba^{-1}b^{-1}\rangle$.

The final things I'll mention here is a construction called the *amalgamated product*.

Definition 2.7.5. Let G, H, K be groups. The free product $G * J$ of two groups is $F_{G,H}$, with simplification by in-group multiplication in G or H also enforced. Let $\varphi \in \mathrm{Hom}(K, G), \psi \in \mathrm{Hom}(K, H)$. Then the amalgamated product of G, H over K with respect to φ, ψ, which is denoted $G *_K H$, is given by

$$G *_K H = \frac{G * H}{\langle\{\varphi(k)\psi(k)^{-1}\}_{k\in K}\rangle_N}$$

All of these constructions are quite important in algebraic topology, but won't be used much by us for the remainder of this text.

2.8 Sylow's Theorems

This section focuses on Sylow's theorems, an important tool in classifying and understanding the structure of finite groups. The structure and proofs in the section are primarily based on those in [Lan05], with some inspiration taken from [Jac09]. We open, however, with a quick detour to talk about exponents.

Definition 2.8.1. Let G be a group. The *exponent* of G, denoted $\exp(G)$, is the minimal number $n \in \mathbb{N}$ such that $g^n = 1$ for all $g \in G$. If no such number exists, we write that $\exp(G) = \infty$.

Lemma 2.8.2. *Let G be a group with $\exp(G) = n < \infty$. Then given any $g \in G$, $o(g) \mid n$.*

Proof. By definition, $o(g) \leq n$. Suppose that $o(g) \nmid n$. Then $\exists k \in \mathbb{N}, 0 < \ell < o(g)$ such that $n = ko(g) + \ell$. Thus, since $g^n = 1$, we get that $g^\ell = 1$. But $\ell < o(g)$ is non-zero, so this is impossible. □

Lemma 2.8.3. *Let G be a finite Abelian group with $\exp(G) = n < \infty$. Then $|G| \mid n^k$ for some $k \in \mathbb{N}$.*

Proof. We proceed by induction on $|G|$, assuming that $\exp(G) = n$. If $|G| = \exp(G) = n$, then the result is trivial. Suppose that the result holds for all groups G with $n \leq |G| < m$. Let G be a group of size m, with $\exp(G) = n$. Pick any non-trivial $h \in G$. By lemma 2.8.2, $o(h) \mid n$, so $H = \langle h\rangle$ has an order which divides n. Since G is Abelian, $H \trianglelefteq G$, so G/H is again an Abelian group and $\exp(G/H) \mid n$. Thus, by induction we get that $|G/H| \mid n^k$ for some $k \in \mathbb{N}$, so by Lagrange's theorem $m = [G : H]|H| \mid n^{k+1}$, completing the proof. □

Using these results, we can start to work on Sylow's theorems.

Definition 2.8.4. Let G be a finite group, and p a prime which divides $|G|$. A subgroup $H \subseteq G$ is called a p-subgroup if $|H| = p^n$ for some $n \in \mathbb{N}$, and a p-Sylow subgroup if this n is the maximal natural number such that $n \mid |G|$.

Lemma 2.8.5. *Let G be a finite Abelian group, and p a prime such that $p \mid |G|$. Then there exists an element of G of order p.*

Proof. Again, we proceed by induction on $|G|$. The result is clear for $|G| \leq p$. Suppose it holds for all $|G| < n$, where $p < n$, and that $|G| = n$. By lemma 2.8.3, $n \mid \exp(G)^k$ for some $k \in \mathbb{N}$, so $p \mid \exp(G)$. Since $\operatorname{lcm}\{o(g) \mid g \in G\} \mid \exp(G)$, it follows that there exists some $g \in G$ such that $p \mid o(g)$. If G is cylic, the result is then clear, and otherwise the result then holds by induction on $H = \langle g \rangle$. □

Theorem 2.8.6 (Sylow I)**.** *Let G be a finite group, and p a prime number dividing the order of G. Then there exists a subgroup $H \subseteq G$ of order p^k for each $k \in \mathbb{N}$ such that $p^k \mid |G|$. In particular, G has a p-Sylow subgroup.*

Proof. Again, we proceed by induction on $|G|$. If $|G| \leq p$ then the result is trivial. Suppose the result holds for all $|G| < n$, where $n > p$, and that $|G| = n$. If there exists some subgroup $H \subsetneq G$ such that $p \nmid [G : H]$, then the result is immediate by induction, since $|H| < |G|$. Thus, we may assume that $p \mid [G : H]$ for all subgroups $H \subsetneq G$. From the class equation, we know that

$$n = |C(G)| + \sum_{i=1}^{n} [G : C(x_i)]$$

where the x_i are representatives of the non-trivial conjugacy classes of G. Since $p \mid [G : C(x_i)]$ for each x_i, it follows that $p \mid |C(G)|$, that is G has a non-trivial centre. Then by lemma 2.8.5, there exists an element $g \in C(G)$ of order p. If $p^k \mid |G|$ only for $k = 1$, we're done. Otherwise, we note that $H = \langle g \rangle \trianglelefteq G$, so G/H is a group of order $\frac{|G|}{p}$. Let N be the maximal number such that $p^N \mid |G|$ By induction, we can find a subgroup $K' \in G/H$ of order p^k for $1 \leq k < N$. Let $\varphi : G \to G/H$ be the canonical quotient map, and $K' = \varphi^{-1}(K)$. Then by the first fundamental theorem of homomorphisms $\frac{K'}{H} \cong K$, so by Lagrange's theorem $|K'| = |K||H| = p^{k+1}$, completing the proof. □

To prove the second Sylow theorem, we need one more lemma.

Lemma 2.8.7. *Let G be a p-group acting on a finite set X. Then the number of $x \in X$ such that $Stab_G(x) = G$ is equivalent to $|X|$ modulo p.*

Proof. Let S be the set of $x \in X$ such that $\operatorname{Stab}_G(x) = G$. Then by Theorem 2.6.7

$$|X| = |S| + \sum_{i=1}^{n} [G : \operatorname{Stab}(x_i)]$$

where the x_i are representative of all the other G-orbits. Since G is a p-group, $p \mid [G : \operatorname{Stab}(x_i)]$ for each x_i, which immediately gives the desired result. □

Theorem 2.8.8 (Sylow's Theorem II)**.** *Let G be a finite group, and p a prime such that $p \mid |G|$. Then the following all hold.*

1. *Every p-subgroup of G is contained in a p-Sylow subgroup of G.*
2. *Every p-Sylow subgroup of G is conjugate, that is if P_1, P_2 are p-Sylow subgroups then $\exists g \in G$ such that $gP_1g^{-1} = P_2$.*
3. *The number of p-Sylow subgroups of G is equivalent to one modulo p.*

Proof. Consider the action of G on the set Γ of p-Sylow subgroups by conjugation. We give a special name to $\mathrm{Stab}_G(P)$, where $P \in \Gamma$; we call it the normalizer of P and denote it $N(P)$. Let H be a p-subgroup. First, let's suppose that $H \subseteq N(P)$. Then $HP \subseteq N(P)$ and $P \trianglelefteq HP$, so by the third fundamental theorem of homomorphisms

$$\frac{HP}{P} \cong \frac{H}{H \cap P}$$

Thus, if $HP \neq P$ then HP is a p-subgroup with order larger than P, which is impossible. Hence, $HP = P \Rightarrow H \subseteq P$, as was required for (1). We move now to a more general case. Let H act on the set of all conjugate subgroups of P, called S, by conjugation. We can see that $|S| = \frac{|N(P)|}{|P|}$, so $p \nmid |S|$. Thus, it follows by lemma 2.8.7 that there exists at least one element of $gPg^{-1} \in S$ which is unchanged by any element of H, meaning $H \subseteq gPg^{-1}$ and hence by the previous case $H \subseteq gPg^{-1}$. This completes the proof of (1). If we take H to be a p-*Sylow* subgroup, then $|H| = |P|$, so we instead get that $H = gPg^{-1}$ for some $g \in G$, giving (2). For (3), we note that since a general p-subgroup $H \subset gPg^{-1}$ is contained in some p-Sylow subgroup, and so no other p-Sylow subgroup is unchanged by every element of H under conjugation. Therefore, there is exactly one element of S has $\mathrm{Stab}_G(x) = G$, so by lemma 2.8.7 we get $|S| \equiv 1 \bmod p$, proving (3). □

We'll end this section off by taking a look at a simple yet very powerful application of Sylow's theorems : the structure theorem of finite Abelian groups. First, we need to intoduce the concept of direct products.

Definition 2.8.9. Let G_1, G_2 be two groups. The *direct product* of the groups, denoted $G_1 \times G_2$ is the group whose set is $G_1 \times G_2$ (the Cartesian product) and operation element-wise multiplication in the two groups.

We'll leave it to the reader to show that this is indeed a group, and that the direct product of groups is associative and commutative up to isomorphism. Indeed, taking the direct product of an arbitrary number of groups is simply taking the Cartesian product of those groups with element-wise multiplication. Before moving on, we do need one more result on direct products.

Theorem 2.8.10. *Let G be a group and $\{H_i\}_{i=1}^n$ a finite collection of subgroups of G. Then $G \cong H_1 \times \cdots \times H_n$ if the following two conditions are met.*

1. *Every element of G can be expressed uniquely in the form $h_1 \cdots h_n$, where $h_i \in H_i$*

2. Every element of H_i commutes with every element of H_j, for all $1 \leq i, j \leq n, i \neq j$

Proof. Suppose $\{H_i\}_{i=1}^n$ is a collection of subgroups meeting the above conditions. The first condition gives us a natural bijective map $\varphi : H_1 \times \cdots \times H_n \to G$ defined by $\varphi : (h_1, \ldots, h_n) \mapsto h_1 \cdots h_n$. All that remains is to check that this is a homomorphism. Let $(h_1, \ldots, h_n), (h_1', \ldots, h_n') \in H_1 \times \cdots \times H_n$. Then

$$\varphi((h_1, \ldots, h_n)(h_1', \ldots, h_n')) = \varphi((h_1h_1', \ldots, h_nh_n')) = (h_1h_1') \cdots (h_nh_n')$$

By condition (2), all the elements on the right-hand side commute, so

$$\varphi((h_1, \ldots, h_n)(h_1', \ldots, h_n')) = (h_1h_2 \cdots h_n)(h_1' \cdots h_n') = \varphi((h_1, \ldots, h_n))\varphi((h_1', \ldots, h_n'))$$

as required. □

Theorem 2.8.11 (Structure Theorem of Finite Abelian Groups)**.** *Let G be a finite Abelian group. Then there exist some $p_i, e_i \in \mathbb{N}$ such that*

$$G \cong \mathbb{Z}/p_1^{e_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p_n^{e^n}\mathbb{Z}$$

where each p_i is a prime such that $p_i^{e_i} \mid |G|$. Furthermore, this decomposition is unique.

Proof. We start by proving existence. First, suppose that $|G| = p^n$, for some prime p. We proceed by induction on n. Suppose that the theorem holds for all $|G| = p^m$, where $m < n$. Let $h \in G$ be an element of maximal order $N \in \mathbb{N}$, and $H = \langle h \rangle$. We have by induction that $G/H \cong \langle g_1H \rangle \times \cdots \times \langle g_rH \rangle$, where $g_i \in G \backslash H$. We'll first show that we can choose $g_i \in \langle g_iH \rangle$ such that $o(g_i) = |\langle g_iH \rangle|$. We know that there exists some minimal $\ell, q \in \mathbb{N}$ such that $g_i^{p^\ell} = h^{zp^q}$, where $z \in \mathbb{N}$ and $1 \leq z < p$. Furthermore, $o(g_i) = p^{\ell+N-q}$, and by the maximality of N we know that $q \geq 1$. We want for $o(g_i) = p^\ell$. This is simply achieved by multiplying g_i by $h^{-zp^{q-\ell}}$, as since g_iH generates $\langle g_iH \rangle$ we know that $o(g_ih) \geq p^\ell$ for any $h \in H$. Thus, we may assume that $o(g_i) = |\langle g_iH \rangle|$. In this case, we define a map $\varphi : H \times G/H \to G$ by the rule

$$\varphi : (h^q, (g_1^{e_1}H, \ldots, g_r^{e_r}H)) \mapsto h^q g_1^{e_1} \cdots g_r^{e_r}$$

We'll show that this is a well-defined isomorphism, which will complete the proof in this case. That this is bijective is immediate via a simple counting argument, since $|\langle g_iH \rangle| = o(g_i)$ and $|G| = |H|[G : H]$. We just then need to show that it's a homomorphism. Indeed, since $o(g_i) = |\langle g_iH \rangle|$ we get

$$\begin{aligned}\varphi((h^q, (g_1^{e_1}H, \ldots, g_r^{e_r}H))(h^\ell, (g_1^{x_1}H, \ldots, g_r^{x_r}H))) &= \varphi((h^{q+\ell}, (g_1^{e_1+x_1}H, \ldots, g_r^{e_r+x_r}H))) \\ = h^{q+\ell}g_1^{e_1+x_1} \cdots g_r^{e_r+x_r} &= \varphi((h^q, (g_1^{e_1}H, \ldots, g_r^{e_r}H)))\varphi((h^\ell, (g_1^{x_1}H, \ldots, g_r^{x_r}H)))\end{aligned}$$

as required.

Next, we prove existence in the general case. Since G is Abelian, we get by Sylow II that there exists a unique p_i-Sylow subgroup P_i of G for each prime $p_i \mid |G|$. By the above case, it suffices to prove that $G \cong P_1 \times \cdots \times P_n$. By Theorem 2.8.10, since G is Abelian it suffices for this to prove that each element of G can be uniquely expressed as a product of one

element from each P_i. Since $\prod_{i=1}^{n} |P_i| = |G|$, uniqueness comes for free with existence of the representation as a product. Again, we can proceed by induction on $|G|$. Suppose this holds for groups of size less than G. Since G is Abelian, $P_1 \trianglelefteq G$, so G/P_1 is an Abelian group with p-Sylow subgroups being the images of $P_2, \dots, P_n$ under the quotient map. Pick any $g \in G$. Then there exists $h \in G \backslash P_1, k \in P_1$ such that $g = hk$. By the inductive hypothesis, there exists $h_2 \in P_2, \dots, h_n \in P_n$ such that $hP_1 = (h_2 \cdots h_n)P_1$. Thus, $\exists h_1 \in P_1$ such that $g = h_1 \cdots h_n$, as required. This completes the proof of existence.

Finally, we prove uniqueness. Suppose $G \cong \mathbb{Z}/p_1^{e_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p_n^{e^n}\mathbb{Z}$, with isomorphism $\varphi : \mathbb{Z}/p_1^{e_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p_n^{e^n}\mathbb{Z} \to G$. By Sylow I, $\varphi(\mathbb{Z}/p_i^{e_i}\mathbb{Z})$ is contained in the p_i Sylow subgroup, so it suffices to show uniqueness of the decomposition of each p_i-Sylow subgroup. Hence, we may assume that $|G| = p^n$ for some prime p. In this case, the result follows from counting the number of distinct subgroups of size $\exp(G)$, then size $\exp(G)/p$, and so on. I'm getting pretty tired at this point since I'm writing this on vacation, so I'll leave the details as an exercise to the reader. □

2.9 Solvable Groups

This section of a combination of similar sections in [Lan05] and [Jac09]. We start with a series of definitions.

Definition 2.9.1. Let G be a group, and $\{G_i\}_{i=1}^{n}$ a collection of subgroups. If $G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_n$, we call this sequence of groups a *tower* of subgroups. A tower of subgroups is normal if $G_{i+1} \trianglelefteq G_i$ for all $0 \leq i < n$, Abelian if it is normal and G_i/G_{i+1} is Abelian for all $0 \leq i < n$, and cyclic if it is normal and G_i/G_{i+1} is cyclic for all $0 \leq i < n$.

Note. We assume that groups are never repeated in a tower. This results in essentially no loss of information, since we can always just toss out the repeated group, and makes proofs a little easier.

At this point, we're already prepared to give the definition of a solvable group.

Definition 2.9.2. A group G is solvable if it has an Abelian tower of subgroups, which terminates with the subgroup $\{1\}$.

Note. The last requirement in this definition is primarily for convenience, as we can always just add $\{1\}$ to the end of any tower that doesn't already have it.

We would also like to be able to have some notion of how "fine" or "course" a normal tower of subgroups is, leading us to the following definition.

Definition 2.9.3. Let G be a group, and $\{G_i\}_{i=1}^{n}$ a normal tower of subgroups. A *refinement* of this tower is any normal tower of subgroups of G which contains $\{G_i\}_{i=1}^{n}$, and is called *proper* if it contains subgroups which are not in the original tower. A normal tower of subgroups is called a *composition series* if $G_n = \{1\}$ and the tower has no proper refinements.

Note. There's another way to characterize composition series. We can note that a normal H subgroup of G is maximal if and only if G/H is *simple*, that is only has normal subgroups of G/H and the trivial group. Thus, a composition series is a normal tower terminating at the trivial group such that each G_i/G_{i+1} is simple.

It turns out that we can simplify our consideration of solvable groups considerably with these refinements. To begin with, we'll need the following lemmas.

Lemma 2.9.4. *Let G, H be groups, $\varphi \in \mathrm{Hom}(G, H)$ a surjective homomorphism, and $G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_n \supseteq \ker(\varphi)$ a tower of subgroups. This tower is normal/Abelian/cyclic if and only if the tower $H \supseteq \varphi(G_1) \supseteq \cdots \supseteq \varphi(G_n)$ is normal/Abelian/cyclic.*

Proof. By the second fundamental theorem of homomorphisms, $G_{i+1} \trianglelefteq G_i$ if and only if $\varphi(G_{i+1}) \trianglelefteq \varphi(G_i)$, for all $0 \leq i < n$, so the normal part of this lemma is clear. Furthermore, we know from this theorem that

$$\frac{G_i}{G_{i+1}} \cong \frac{\varphi(G_i)}{\varphi(G_{i+1})}$$

which gives us the Abelian/cyclic part of the lemma. □

Lemma 2.9.5. *Let G be a solvable group, and $H \trianglelefteq G$. Then G/H is solvable. Furthermore, if G is any group and there exists $H \trianglelefteq G$ such that G/H, H are solvable, then G is solvable.*

Proof. Suppose that G is solvable, and $H \trianglelefteq G$. Since G is solvable, it has an Abelian tower $G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_n = \{1\}$. Let $\pi : G \to G/H$, and set $H_i = \pi(G_i)$. We'll first show that $H_{i+1} \trianglelefteq H_i$. Pick any $g_i \in G_i, g_{i+1} \in G_{i+1}$. Then $(g_iH)^{-1}(g_{i+1}H)(g_iH) = (g_i^{-1}g_{i+1}g_i)H \in G_{i+1}H$, as required. Next, we'll show that there exists a surjective homomorphism from G_i/G_{i+1} to H_i/H_{i+1}, implying the latter is Abelian and proving that G/H is solvable. We'll define this map by $\varphi : gG_{i+1} \mapsto \pi_i(gH)$, where $\pi_i : H_i \to H_i/H_{i+1}$ is the standard projection map. We show that this is a homomorphism first.

$$\varphi((g_1G_{i+1})(g_2G_{i+1})) = \varphi((g_1g_2)G_i) = \pi_i((g_1g_2)H) = \pi_i(g_1H)\pi_i(g_2H) = \varphi(g_1G_{i+1})\varphi(g_2G_{i+1})$$

That it is surjective is immediate. Now, suppose that G is some group and there exists $H \cong G$ such that G/H is solvable. Let $G/H = H_0 \supseteq H_1 \supseteq \cdots \supseteq H_n = \{1\}$ be an Abelian tower, and $\pi : G \to G/H$ the projection map. Then by the second fundamental theorem of homomorphisms, we can get a normal tower of subgroups $G = G_0 = \pi^{-1}(H_0) \supseteq G_1 = \pi^{-1}(H_1) \supseteq \cdots \supseteq G_n = \pi^{-1}(H_n) = H$. Furthermore,

$$\frac{G_i}{G_{i+1}} \cong \frac{H_i}{H_{i+1}}$$

is Abelian, so since H is solvable we're done. □

Theorem 2.9.6. *A finite group G is solvable if and only if it has a cyclic composition series.*

Proof. The having a cyclic composition series implies that a group is solvable is immediate. Now, suppose that G is a solvable finite group. First, we show that G has an Abelian composition series. Since G is solvable, it has an Abelian tower $G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_n =$

$\{1\}$. Any refinement of this tower remains Abelian, so since G is finite we can take a maximal refinement to get an Abelian composition series. Thus, we may assume that our Abelian tower is a composition series. Suppose G_i/G_{i+1} were not cyclic. Then since its Abelian, picking any non-trivial $x \in G_i/G_{i+1}$ we'd get a normal subgroup $G_{i+1} \subsetneq \pi^{-1}(\langle x \rangle) \subsetneq G_i$, where $\pi : G_i \to G_{i+1}$ is the standard projection map. But this contradicts our assumption that the normal tower was a composition series. □

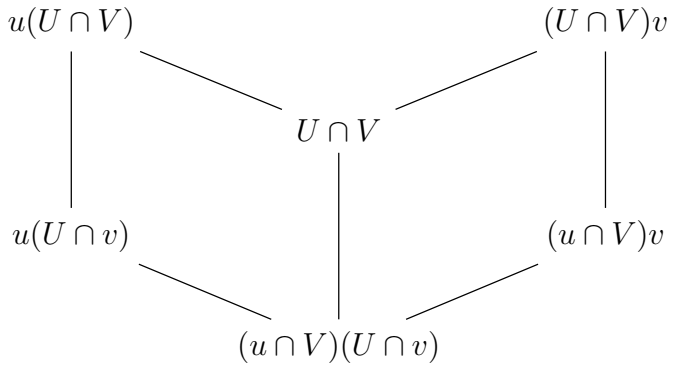
Note. Since cyclic simple groups are necessarily of prime power order, this also implies that the composition series has $G_i/G_{i+1} \cong \mathbb{Z}/p_i\mathbb{Z}$, where p_i are prime.

This can be strengthened even further, there's a notion in which all composition series are equivalent. Thus, a finite would group would be solvable if and only if all of its composition series are cyclic. Let's work towards proving this equivalence now, following [Lan05].

Lemma 2.9.7 (Buttefly Lemma)**.** *Let G be a group, $U, V \subseteq G$ subgroups, and $u \trianglelefteq U, v \trianglelefteq V$ normal subgroups. Then*

1. $u(U \cap v) \trianglelefteq u(U \cap V)$
2. $(u \cap V)v \trianglelefteq (U \cap V)v$
3. $\frac{u(U\cap V)}{u(U\cap v)} \cong \frac{(U\cap V)v}{(u\cap V)v}$

Proof. This is our first instance of proof by pretty picture.



First, we'll show that the intersection of two lines going down is the intersection of two groups, and the intersection of two lines going up is the group generated by the two groups. Clearly $U \cap V \subseteq u(U \cap V) \cap (U \cap V)v$. Suppose $g \in u(U \cap V) \cap (U \cap V)v$. Then $g = xy = wz$, where $x \in u, y, w \in U \cap V, z \in v$. Thus, $x = w(zy^{-1}) = (wzw^{-1})(wy^{-1})$. Since $z \in v$ and $w \in V$, $k = wzw^{-1} \in v$. Since $y^{-1}, w \in U \cap V$, $wy^{-1} = r \in U \cap V$, so $x \in u \cap v(U \cap V) \Rightarrow (v \cap U)(U \cap V) \Rightarrow x \in U \cap V$, and hence $g \in U \cap V$. Therefore, $U \cap V = u(U \cap V) \cap (U \cap V)v$, as claimed. We can again see that $u(U \cap v) \cap (u \cap V)v \supseteq (u \cap V)(U \cap v)$. Suppose $g \in u(U \cap v) \cap (u \cap V)v$. Then $g = xy = wz$, where $x \in u, y \in U \cap v, w \in u \cap V, z \in v$. Thus, $x = wzy^{-1} \in V$, so $x \in u \cap V$ and hence $u(U \cap v) \cap (u \cap V)v = (u \cap V)(U \cap v)$ as claimed. Again, we can see that $(u \cap V)(U \cap v) \subseteq (U \cap V) \cap u(U \cap v)$. Suppose $g \in (U \cap V) \cap u(U \cap v)$. Then $g = xy$, where $x \in u$ and $y \in U \cap v$. Since $g \in U \cap V$, it follows that $x = gy^{-1} \in V$, so $x \in u \cap V$ and

hence $(u \cap V)(U \cap v) = (U \cap V) \cap u(U \cap v)$ as claimed. The remaining case proceeds by an identical argument. The claim about lines going up being generating groups is clear. Second, we can see from the diagram that (1) and (2) are equivalent, so we prove only (1). Suppose $g \in u(U \cap V), h \in u(U \cap v)$. Then $g = xy, h = wz$, where $x, w \in u, y \in U \cap V, z \in U \cap v$. Thus,

$$ghg^{-1} = (xy)(wz)(y^{-1}x^{-1}) = x(ywy^{-1})(yzy^{-1})x^{-1}$$

$r = ywy^{-1} \in u, k = yzy^{-1} \in U \cap v$, so we get

$$ghg^{-1} = xrkx^{-1} = (xr)(kx^{-1}k^{-1})k$$

$xr \in u, kx^{-1}k^{-1} \in u$, so $ghg^{-1} \in u(U \cap v)$ and hence $u(U \cap v) \trianglelefteq u(U \cap V)$, as claimed. Finally, we prove (3). By the third fundamental theorem of homomorphisms

$$\begin{aligned}\frac{u(U \cap V)}{u(U \cap v)} &= \frac{(U \cap V)u(U \cap V)}{u(U \cap v)} \cong \frac{U \cap V}{u(U \cap v) \cap (U \cap V)} = \frac{U \cap V}{(u \cap V)(U \cap v)} \\ &= \frac{U \cap V}{(u \cap V)v \cap (U \cap V)} \cong \frac{(U \cap V)(u \cap V)v}{(u \cap V)v} = \frac{(U \cap V)v}{(u \cap V)v}\end{aligned}$$

as claimed. □

Now, we need to define what we mean when we say two towers are equivalent.

Definition 2.9.8. Let G be a group, and $G = G_1 \supseteq G_2 \supseteq \cdots \supseteq G_n = \{1\}, G = H_1 \supseteq H_2 \supseteq \cdots \supseteq H_m = \{1\}$ be two normal towers. We call these towers *equivalent* if $r = s$ and there exists $\sigma \in S_{n-1}$ such that $\frac{G_i}{G_{i+1}} \cong \frac{H_{\sigma(i)}}{H_{\sigma(i)+1}}$ for all $1 \leq i < n$.

Which will allow us to, finally, develop results on these equivalences.

Theorem 2.9.9 (Schreier's Theorem)**.** *Suppose $G = G_1 \supseteq G_2 \supseteq \cdots \supseteq G_n = \{1\}, G = H_1 \supseteq H_2 \supseteq \cdots \supseteq H_m = \{1\}$ are two normal towers. Then they have equivalent refinements.*

Proof. For each $1 \leq i < n, 1 \leq j \leq m$, define $G_{ij} = G_{i+1}(G_i \cap H_j)$. We can note that $G_{i1} = G_i, G_{im} = G_{i+1}$, so by the butterfly lemma

$$G \supseteq G_{12} \supseteq \cdots \supseteq G_{1m} \supseteq G_{21} \supseteq \cdots \supseteq G_{(n-1)1} \supseteq \cdots \supseteq \{1\}$$

is a refinement of the first normal tower. Similarly, setting $H_{ji} = (H_j \cap G_i)H_{j+1}$ for all $1 \leq i \leq n, 1 \leq j < m$, we get

$$G \supseteq H_{12} \supseteq \cdots \supseteq H_{1n} \supseteq H_{21} \supseteq \cdots \supseteq H_{(m-1)1} \supseteq \cdots \supseteq \{1\}$$

is a refinement of the second normal tower. By the butterfly lemma

$$\frac{G_{ij}}{G_{i(j+1)}} \cong \frac{H_{ji}}{H_{j(i+1)}}$$

for all $1 \leq i < n, 1 \leq j < m$ completing the proof. □

Theorem 2.9.10 (Jordan-Hölder)**.** *If G is solvable, then any two composition series of G are equivalent*

Proof. By Schreier's theorem, the two composition series must have equivalent refinements. But composition series have no proper refinements, so it follows that the composition series are equivalent. □

At this point, we've proven everything I find particularly enlightening (at least at this point in my life) about solvable groups, without of course getting into their major role in Galois theory. We'll end this section by taking a look at the connection between solvability and commutators, following [Jac09].

Definition 2.9.11. Let G be a group and $g, h \in G$. We denote the commutator by

$$[g, h] = g^{-1}h^{-1}gh$$

The *derived* group $G' \subseteq G$ is the subgroup generated by all the commutators of two elements in G. We define the n-th derived group iteratively by $G^{(n)} = (G^{(n-1)})'$.

Lemma 2.9.12. *For any $k \in \mathbb{N}$, $G^{(k)} \trianglelefteq G$.*

Proof. We can note that since $[g, h]^{-1} = h^{-1}g^{-1}hg = [h, g]$, G' consists of elements of the form $[g_1, h_1] \cdots [g_n, h_n]$, where $g_i, h_i \in G$. Let H be another group, and $\varphi \in \mathrm{Hom}(G, H)$. Then $\varphi([g, h]) = [\varphi(g), \varphi(h)] \in H'$, so $\varphi(G') \subseteq H'$. Now, pick any $K \trianglelefteq G$ and $a \in G$. The map $\varphi : g \mapsto aga^{-1}$ induces an automorphism of K, so $\varphi(K') \subseteq K'$. Since a was arbitrary, this implies that $K' \trianglelefteq G$. In particular, $G \trianglelefteq G \Rightarrow G' \trianglelefteq G \Rightarrow \cdots \Rightarrow G^{(k)} \trianglelefteq G$. □

Lemma 2.9.13. *G/G' is Abelian and if $K \trianglelefteq G$ is such that G/K is Abelian then $G' \subseteq K$.*

Proof. Let $\pi \in \mathrm{Hom}(G, G/G')$ be the projection map, and pick any $g, h \in G$. Then $\pi(g)\pi(h) \equiv g(g^{-1}hgh^{-1})h \equiv hg \equiv \pi(h)\pi(g)$, so G/G' is Abelian. Now, suppose that $K \trianglelefteq G$ is such that G/K is Abelian, and pick any $g, h \in G$. Then

$$(g^{-1}h^{-1}gh)K = (gK)^{-1}(hK)^{-1}(gK)(hK) = (gK)^{-1}(gK)(hK)^{-1}(hK) = K$$

so $g^{-1}h^{-1}gh \in K \Rightarrow G' \subseteq K$, as was to be shown. □

Theorem 2.9.14. *A group G is solvable if and only if there exists $n \in \mathbb{N}$ such that $G^{(n)} = \{1\}$.*

Proof. One direction is simple, as $G \supseteq G' \supseteq \cdots \supseteq G^{(n)}$ is an Abelian tower by lemmas 2.9.12 and 2.9.13. Now, suppose that G is solvable, and that $G = H_0 \supseteq H_1 \supseteq \cdots \supseteq H_k = \{1\}$ is an Abelian tower. By lemma 2.9.13, $G' \subseteq H_1$, so $G'' \subseteq H_1' \subseteq H_2$ and so on, forcing $G^{(k)} = \{1\}$ as desired. □

2.10 Group Representations*

As you may have surmised from the past chapter, understanding the structure of groups can be quiet difficult. One of the ways of getting around this is group representations, which we'll introduce (but not develop all that much) here. The study of representations is a field unto itself; those interested should look at [Lan05] or one of the numerous textbooks dedicated to the subject.

Given a vector space V, one can see that the set of all invertible linear maps from V to V, with composition as multiplication, forms a group. Since we understand linear transformations far more than we do groups, our aim is to shift the study of groups to the study of linear algebra.

Definition 2.10.1. Let G be a group and V a vector space. A representation of G in V is the image of a homomorphism from G to the invertible linear transformations of V to itself.

Example 2.10.1. Given any group G and vector space V, we have the trivial representation given by $\varphi : g \mapsto \mathrm{Id}_V$. Needless to say this one is not particularly useful.

Example 2.10.2. Consider the group $\Gamma = \{e^{i\frac{2\pi j}{n}}\}_{1\leq j\leq n}$, where the operation is multiplication. We can represent this in $\mathrm{GL}_2(\mathbb{C})$ using the map

$$\varphi : e^{i\frac{2\pi j}{n}} \mapsto \begin{pmatrix} 1 & 0 \\ 0 & e^{i\frac{2\pi j}{n}} \end{pmatrix}$$

There are also many ways of classifying representations, we introduce some of them below.

Definition 2.10.2. Let G be a group represented in a vector space V, with the associated homomorphism being φ. A representation is said to *irreducible* if, given any non-trivial subspace $U \subsetneq V$, there exists some $\underline{v} \in U$ and $g \in G$ such that $\varphi(g)(\underline{v}) \notin U$. Otherwise, it is called *reducible*. A representation is called *faithful* if φ is injective.

Chapter 3

Rings

3.1 Basic Definitions

When studying group theory, two of our most common examples were the monoids $(\mathbb{Z}, +)$ and $(\mathbb{Z}^*, \cdot)$ (non-zero integers under multiplication). Of course, there is something rather unnatural about treating these two monoids as separate objects; we know intuitively that they are parts of a single object, the integers. Our resolution to this is the ring.

Definition 3.1.1. A ring R is a set with elements $0, 1 \in R$ (which we call the zero and identity) and binary operations $+, \cdot$ such that

1. $(R, +, 0)$ is an Abelian group.
2. $(R, \cdot, 1)$ is a monoid.
3. For any $x, y, z \in R$, distributivity is respected. That is,

$$(x + y) \cdot z = x \cdot z + y \cdot z$$
$$z \cdot (x + y) = z \cdot x + z \cdot y$$

Note. Like with groups, we often drop the $\cdot$ when writing products. We will often denote $R\backslash\{0\}$ by R^*. The definition of a ring will vary from text to text, some older sources do not assume that all rings have an identity. Others call rings without an identity rngs (this is a ploy used by [Jac09]). Either way, it will not be of much interest to us here. All of our rings, by assumption, will have an identity.

Example 3.1.1. Perhaps the best example of a ring is $M_n(\mathbb{R})$, the set of $n \times n$ real matrices with matrix addition and multiplication. Our identity here is the identity matrix, and our zero the zero matrix. Note that multiplication here is **not** commutative: assuming it to be so is a common mistake when working with rings.

There are many types of properties a ring can have, we list them here.

Definition 3.1.2. A ring R such that $0 \neq 1$ is

1. Commutative if $(R, \cdot, 1)$ is Abelian.

2. A (integral) domain if $(R^*, \cdot, 1)$ is a sub-monoid.

3. A division ring if $(R^*, \cdot, 1)$ is a sub-group.

4. A field if it is a commutative division ring.

Note. The assumption that $0 \neq 1$ here is very common, so much so that it will often be assumed without being stated.

We'll come back to all of these in a moment to explore their connections more deeply. For now, we note some basic results on arithmetic in rings.

Proposition 3.1.3. *Let R be a ring, and $x, y \in R$ ring elements. Then*

1. $1x = x1 = x$

2. $0x = x0 = 0$

3. $(-1)x = -x$

4. *If $xy = yx$, then for any $n, m \in \mathbb{N}$ we get $x^n y^m = y^m x^n$, and*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

The proof of these basic identities is left to the reader. Like with groups, we of course have subrings and generated subrings.

Definition 3.1.4. A subring $S \subseteq R$ of a ring R is a subset which is also a ring. For an arbitrary subset $A \subset R$, then subring generated by A, $\langle A \rangle$, is the smallest subring of R containing A.

Like with group and monoids, the intersection of subrings is a subring, and hence $\langle A \rangle$ is the intersection of all subrings of R containing A. The elements of $\langle A \rangle$ are $0, 1$, and all finite sums of finite products of elements of A. With these basic concepts out of the way, we return to the problem of characterizing the properties of rings.

Definition 3.1.5. If $a \in R$ is such that there exists some non-zero $b \in R$ for which $ab = 0$ ($ba = 0$), then a is called a left (right) zero-divisor of b.

Theorem 3.1.6. *The following are equivalent (where R is a ring and $R \neq 0$).*

1. *R is a domain.*

2. *R has no non-zero zero-divisors.*

3. *For any $x, y, z \in R$, $xy = xz \Rightarrow y = z$ or $x = 0$ and $xy = zy \Rightarrow x = z$ or $y = 0$. This condition is called the cancellation law.*

Proof. First, suppose that R is a domain. Then since R^* is a sub-monoid, the product of any pair of non-zero elements is non-zero, and hence R has only 0 as a zero divisor. Now, suppose that R has no non-zero zero-divisors. If $xy = xz$, then $x(y - z) = 0$, so either $x = 0$ or $y - z = 0 \Rightarrow y = z$. If $xy = zy$, then $(x - z)y = 0$, so either $y = 0$ or $x - z = 0 \Rightarrow x = z$. Finally, suppose that the cancellation law holds, and pick any pair of non-zero elements $x, y \in R$. If $xy = 0$, then $xy = x0 \Rightarrow x = 0$ or $y = 0$, a contradiction. Thus, $xy \in R^*$, making R^* a sub-monoid. □

There are two final objects to define before we move on.

Definition 3.1.7. The subgroup of $(R^*, \cdot, 1)$ consisting of elements with a multiplicative inverse is called the units of R, and denoted

$$R^\times = \{x \in R \mid \exists y \in R, xy = yx = 1\}$$

Note. We again denote the multiplicative inverse of $x \in R$ by x^{-1}.

Definition 3.1.8. Let R, R' be rings. A ring homomorphism is a map $\varphi : R \to R'$ which is a group homomorphism $(R, 0, +) \to (R', 0', +')$ and a monoid homomorphism $(R, 1, \cdot) \to (R', 1', \cdot')$. The set of all ring homomorphisms between two rings is denoted $\mathrm{Hom}(R, R')$

Note. By definition, we must get $\varphi(0) = 0'$ and $\varphi(1) = 1'$.

3.2 Matrix Rings

We start with one of the simplest yet most important types of rings, the matrix ring. This section is based on a similar one in [Jac09].

Definition 3.2.1. Let R be an arbitrary ring, and $n \in \mathbb{N}$. The matrix ring of R, denoted $M_n(R)$, is the set of all $n \times n$ matrices with entries in R. Endowed with standard matrix addition and multiplication, $M_n(R)$ is a ring.

Note. We can embed R in $M_n(R)$ via the monomorphism $x \mapsto \mathrm{diag}(x, x \ldots, x)$. Thus, if $A \in M_n(R)$, by xA we mean $\mathrm{diag}(x, x, \ldots, x)A$. We can pull a similar trick and map $\mathbb{Z}$ into any ring R, with the map $f \in \mathrm{Hom}(\mathbb{Z}, R)$ being defined by $f(1) = 1$. In this notation, for any given $n \in \mathbb{Z}$ we write nx for $f(n)x$.

Note. Even if R is commutative, $M_n(R)$ will be non-commutative if $n \geq 2$.

We denote by e_{ij} the matrix with entries all zero except in the i, jth position. Note then that

$$(a_{ij}) = \sum_{i,j} a_{ij} e_{ij}$$

I'll cut right to the chase, let's figure out determinants shall we?

Theorem 3.2.2. *For any commutative ring R, there exists a unique function $f : M_n(R) \to R$ such that*

1. $f(\mathrm{Id}_n) = 1$.
2. *If A' is the matrix A with its rows permuted by some $\sigma \in S_n$, then $f(A') = \mathrm{sgn}(\sigma) f(A)$.*
3. *f is linear in each row of $M_n(R)$, keeping all other rows fixed.*

Proof. Suppose such a function f existed. Pick an arbitrary matrix $A = (a_{ij})$. Then

$$\begin{aligned} f(A) = f\Big(\sum_{i,j} a_{ij} e_{ij}\Big) &= \sum_{k=1}^{n} a_{1k} f\Big(e_{1k} + \sum_{i=2}^{n}\sum_{j=1}^{n} a_{ij} e_{ij}\Big) \\ &= \sum_{k_1=1}^{n} \cdots \sum_{k_n=1}^{n} a_{1k_1} \cdots a_{nk_n} f(e_{1k_1} + \cdots + e_{nk_n}) \end{aligned}$$

Note that if some $k_i = k_j$ in any given sum, then there's a permutation of the rows of $e_{1k_1} + \cdots + e_{nk_n}$ of sign one which doesn't change the matrix. That is, we get

$$f(e_{1k_1} + \cdots + e_{nk_n}) = -f(e_{1k_1} + \cdots + e_{nk_n})$$

Thus, all the terms with $k_i = k_j$ for some $1 \leq i < j \leq n$ cancel out to zero, and we're left with

$$f(A) = \sum_{\sigma \in S_n} a_{1\sigma(1)} \cdots a_{n\sigma(n)} f(e_{1\sigma(1)} + \cdots + e_{n\sigma(n)})$$

But of course each $e_{1\sigma(1)} + \cdots + e_{n\sigma(n)}$ is just the identity with its rows permuted by σ, so

$$f(A) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)} f(\mathrm{Id}_n) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)}$$

Therefore, f is unique if it is well-defined. We just need to check then that the function

$$f(A) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)}$$

satisfies all three properties. This verification is not too difficult, and left to the reader. □

We call this unique function the determinant, and denote it det. This is of course the same determinant as you would have encountered in linear algebra, although perhaps it doesn't seem that way for now. Let's look at different ways of formulating this same function.

Definition 3.2.3. Let $A \in M_n(R)$. The i,jth minor of A, denoted $M_{A,i,j}$, is the determinant of the matrix obtained be removing the ith row and jth column of A, multiplied by $(-1)^{i+j}$.

This definition leads to the determinant which you may be more familiar with. Before that though, we need a quick lemma.

Proposition 3.2.4. *If R is a commutative ring, then properties (2) and (3) of* det *also hold for columns.*

Note. This hints at the fact that our choice to define the determinant properties in terms of columns in Theorem 3.2.2 was arbitrary. In fact, it would be entirely equivalent to phrase Theorem 3.2.2 in terms of columns, and then re-phrase and prove the lemma 3.2.4 in terms of rows.

Corollary 3.2.4.1. *If R is commutative, then for any $A \in M_n(R)$ and $1 \leq i \leq n$*

$$\det(A) = \sum_{j=1}^{n} a_{ij} M_{A,i,j}$$

and

$$\det(A) = \sum_{j=1}^{n} a_{ji} M_{A,j,i}$$

Proof. We note that

$$\det(A) = \sum_{j=1}^{n} a_{ij} \det\left(e_{ij} + A - \sum_{k \neq j} e_{ik}\right)$$

Let's examine $e_{ij} + A - \sum_{k \neq j} e_{ik}$ more closely. Each of these is a matrix of the form

$$\begin{pmatrix} a_{11} & \cdots & a_{1(j-1)} & a_{1j} & a_{1(j+1)} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{(i-1)1} & \cdots & a_{(i-1)(j-1)} & a_{(i-1)j} & a_{(i-1)(j+1)} & \cdots & a_{(i-1)n} \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ a_{(i+1)1} & \cdots & a_{(i+1)(j-1)} & a_{(i+1)j} & a_{(i+1)(j+1)} & \cdots & a_{(i+1)n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{n(j-1)} & a_{nj} & a_{n(j+1)} & \cdots & a_{nn} \end{pmatrix}$$

We can apply a row and column swap to this to end up with the matrix

$$\begin{pmatrix} 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{(i-1)j} & \cdots & a_{(i-1)(j-1)} & a_{(i-1)1} & a_{(i-1)(j+1)} & \cdots & a_{(i-1)n} \\ a_{1j} & \cdots & a_{1(j-1)} & a_{11} & a_{1(j+1)} & \cdots & a_{1n} \\ a_{(i+1)j} & \cdots & a_{(i+1)(j-1)} & a_{(i+1)1} & a_{(i+1)(j+1)} & \cdots & a_{(i+1)n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{nj} & \cdots & a_{n(j-1)} & a_{n1} & a_{n(j+1)} & \cdots & a_{nn} \end{pmatrix}$$

It's clear from the permutation definition of the determinant (and the row/column properties it satisfies) then that $\det\left(e_{ij} + A - \sum_{k \neq j} e_{ik}\right)$ is $(-1)^{i+j}$ multiplied by the determinant of the (1,1) minor of the above matrix. But of course that's just $M_{A,i,j}$, giving us the desired result. The result for column expansion is proved in an essentially identical manner. □

The determinant has one more relevant property, the proof of which we omit[1].

Proposition 3.2.5. $\det : M_n(R) \to R$ *is a monoid homomorphism with respect to matrix multiplication.*

Note. This is just a very fancy way of saying $\det(AB) = \det(A)\det(B)$. The point of viewing it in the above manner is that we can then embed R into $M_n(R)$, giving us a natural map $\det : M_n(R) \to M_n(R)$ defined by (abusing notation a bit) $\det(A) = \text{diag}(\det(A), \ldots, \det(A))$.

We need one more result before we can cover the main application of the determinant.

Lemma 3.2.6. *Suppose R is commutative, $A \in M_n(R)$, and $1 \leq i, j \leq n$ are such that $i \neq j$. Then*

$$\sum_{k=1}^{n} a_{ik} M_{A,j,k} = 0$$

$$\sum_{k=1}^{n} a_{ki} M_{A,k,j} = 0$$

Proof. We prove only the first identity, the proof for the second is essentially the same. Note that by corollary 3.2.4.1, the first expression is just the determinant of the matrix A with its jth row replaced by its ith row. Since $i \neq j$, such a matrix has a repeated row, which by property (2) of determinants implies it has determinant zero. □

Let us now cover the aforementioned application of determinants, invertibility and adjugate matrices.

Definition 3.2.7. Let $A \in M_n(R)$, where R is commutative. The cofactor matrix of A, denoted $\text{adj}(A)$, is the $n \times n$ matrix who's (i, j)th entry is $M_{A,j,i}$.

Note. [Jac09] calls this the adjoint matrix, which is terrible form and should not be done.

Theorem 3.2.8. *Suppose R is commutative and $A \in M_n(R)$. Then A is invertible if and only if* $\det(A)$ *is a unit, and the inverse of A (if it exists) is* $\det(A)^{-1} adj(A)$.

Note. In this statement, and its proof, we use $\det(A)$ to refer both to the determinant of a matrix and the embedding of that determinant back into $M_n(R)$.

Proof. First, suppose that A is invertible. Then by proposition 3.2.5 we get $1 = \det(AA^{-1}) = \det(A)\det(A^{-1})$, so $\det(A)$ must be a unit. Now, suppose that $\det(A)$ is a unit. For any $1 \leq i, j \leq n$, we get

$$(A\text{adj}(A))_{ij} = \sum_{k=1}^{n} a_{ik}(\text{adj}(A))_{kj} = \sum_{k=1}^{n} a_{ik} M_{A,j,k}$$

Thus, by lemma 3.2.6 and corollary 3.2.4.1 we conclude that $A\text{adj}(A) = \det(A)$. A similar calculation implies that $\text{adj}(A)A = \det(A)$, so since $\det(A)$ is a unit we get $A^{-1} = \det(A)^{-1}\text{adj}(A)$, as required. □

In particular, this is just the familiar result from an introductory linear algebra course.

Corollary 3.2.8.1. *If F is a field, then $A \in M_n(F)$ is invertible if and only if* $\det(A) \neq 0$.

[1] The proof is basically symbol pushing

3.3 Ideals and Quotient Rings

As will become a common theme in this text, we wish now to replicate the homomorphism theorems for groups in the setting of rings. In order to do this, we need to figure out how to construct quotient rings. To that end, let us consider a ring R with subring A. Ideally, we would like for the following equations to hold in R/A, for any $x, y \in R$

$$(x + A) + (y + A) = (x + y) + A, (x + A)(y + A) = xy + A$$

The first of these we get for free, in particular it's just a manifestation of the additive group in the ring being Abelian, and all subgroups of Abelian groups being normal. The second is not at all guaranteed, leading us to the following definition.

Definition 3.3.1. A left (right) ideal $I \subseteq R$ is a sub-ring of R such that $RI \subset I$ ($IR \subset I$). A subset which is both a left and right ideal is simply called an ideal.

Note. R is necessarily an ideal of R.

Ideals are in fact the structure we need to generate quotient rings. Indeed, one can expand[2] to get

$$(x + A)(y + A) = xy + xA + Ay + A^2$$

so we need a guarantee that $xA, Ay \subset A$, which is exactly to say that A is an ideal.

Definition 3.3.2. Let R be a ring and $I \subset R$ an ideal. The quotient ring R/I is the quotient group with multiplication defined by, for any $x, y \in R$

$$(x + I)(y + I) = xy + I$$

Again, that this is a well-defined ring follows from the properties of an ideal. That being said, let us explore more properties of ideals.

Proposition 3.3.3. *Let R be a ring, and $\{I_j\}_{j\in J}$ a collection of ideals in R. Then $\bigcap_{j\in J} I_j \subset R$ is an ideal.*

Proof. That it's an additive subgroup follows from Theorem 2.1.5. We just need to check then closure under multiplication. Pick any $a \in \bigcap_{j\in J} I_j$ and $x \in R$. Then for each I_j, $a \in I_j$, and hence $xa, ax \in I_j$. It follows that $xa, ax \in \bigcap_{j\in J} I_j$, as required. □

This, along with the above note about R being an ideal of itself, allows us to define the sub-ring generated by a set.

Definition 3.3.4. Let R be a ring, and $S \subset R$. The ideal generated by S, denoted (S), is the smallest ideal in R containing S (i.e. the intersection of all ideals in R containing S).

Like with rings and monoids, we can explicitly write out the elements of this ideal.

[2]Strictly speaking this is not a "proper" expansion, but you get the idea.

Proposition 3.3.5. *Let R be a ring, and $S \subset R$. Then*

$$(S) = \Big\{ \sum_{a \in S} x_a a y_a \mid x_a, y_a \in R \Big\}$$

where all the above sums have finitely many non-zero terms.

Note. The above proposition and definition have fairly immediate equivalent formulations for right and left ideals.

The proof of this is left to the reader.

Note. I'm going to be doing a lot more "left to the reader" or "it is clear" explanations from now on. The hope is that the rigour of the previous section has given you the intuition to follow and be comfortable with such explanations.

3.4 Homomorphism Theorems

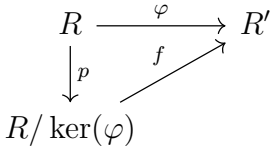
The theorems so nice we cover them twice. These are the theorems presented in [Jac09], although the order and proofs have been changed. We start with a quick lemma.

Lemma 3.4.1. *Let $\varphi \in \mathrm{Hom}(R, R')$, where R, R' are rings. Then $\ker(\varphi) \subset R$ is an ideal.*

Proof. That it's an additive subgroup is immediate by a similar result on group homomorphisms, so it suffices to show that $R\ker(\varphi), \ker(\varphi)R \subset \ker(\varphi)$. To that end, pick any $r \in R, x \in \ker(\varphi)$. Then $\varphi(rx) = \varphi(r)\varphi(x) = 0\varphi(x) = 0$, and $\varphi(xr) = \varphi(x)\varphi(r) = \varphi(x)0 = 0$, so $rx, xr \in \ker(\varphi)$, as required. □

No point in wasting time, let's jump right into these.

Theorem 3.4.2 (First Fundamental Theorem of Homomorphisms)**.** *Let $\varphi : R \to R'$ be a ring homomorphism. Then the natural projection map $p : R \mapsto R/\ker(\varphi)$ is a ring homomorphism, and the map $f : R/\ker(\varphi) \to \mathrm{Im}(\varphi)$ given by $f : x + \ker(\varphi) \mapsto \varphi(x)$ is a well-defined ring isomorphism. Finally, the following diagram commutes.*



Proof. First, pick any $x, y \in R$. Then $p(xy) = xy + I = (x + I)(y + I)$ (here we're doing arithmetic with cosets), and $p(x + y) = (x + I) + (y + I) = (x + y) + I$, making p a ring homomorphism. f is certainly a homomorphism if well-defined (by an essentially identical check), so we check instead that it is indeed well-defined. Pick any $x, y \in R$ such that $x + \ker(\varphi) = y + \ker(\varphi)$. Then $\exists z \in \ker(\varphi)$ such that $x = y + z$. Thus, $\varphi(x) = \varphi(y) + \varphi(z) = \varphi(y)$, making f well-defined. Finally, we show that f is an isomorphism onto $\mathrm{Im}(\varphi)$ (the above diagram commuting is immediate from this). But this is immediate from our verification of f being well-defined. □

Theorem 3.4.3 (Second Fundamental Theorem of Homomorphisms)**.** *Suppose $\varphi : R \to R'$ is a surjective ring homomorphism. Then*

1. *An additive subgroup $S \subset R$ containing $\ker(\varphi)$ is a subring (ideal) of R if and only if $\varphi(S)$ is a subring (ideal) of R'.*
2. *The map $S \mapsto \varphi(S)$ on subrings (ideals) of R containing $\ker(\varphi)$ is a bijection onto subrings (ideals) of R'.*
3. *If $I \subset R$ is an ideal containing $\ker(\varphi)$, then $R/I \cong R'/\varphi(I)$.*

Proof. We start with the first statement. If S is a subring (ideal), then since φ is surjective it is immediate that $\varphi(S)$ is a subring (ideal). Now, suppose that $\varphi(S)$ is a subring. Pick any $x, y \in S$. Then $\varphi(xy) = \varphi(x)\varphi(y) \in \varphi(S)$, so it follows that there exists some $s \in S, z \in \ker(\varphi)$ such that $xy = s + z$. But of course $\ker(\varphi) \subset S$, so this implies that $xy \in S$, and hence S is a subring as it is an additive subgroup. Suppose further that $\varphi(S)$ is an ideal. Pick any $x \in S, r \in R$. Then $\varphi(xr) = \varphi(x)\varphi(r) \in \varphi(S)$, and similar with $\varphi(rx)$. Thus, both differ from an element of S only by some element of $\ker(\varphi)$, which again means that $xr, rx \in S$ and hence S is an ideal.

Now for the second statement. Suppose that S, S' are two subrings (ideals) of R containing $\ker(\varphi)$. Then $\varphi(S) = \varphi(S')$ implies that any element of S not in S' differs only by addition of an element in $\ker(\varphi)$, and vice-versa. But of course both subrings (ideals) contain $\ker(\varphi)$, so this implies that $S = S'$, and hence $S \mapsto \varphi(S)$ is injective. Now, suppose that $S' \subset R'$ is a subring (ideal). It suffices to show that $\varphi^{-1}(S')$ is a subring (ideal). But of course φ is an additive group homomorphism and S' an additive subgroup, so $\varphi^{-1}(S')$ is an additive subgroup of R and our result follows from part (1) of this theorem.

Finally, we prove the third statement. Suppose that $I \subset R$ is an ideal containing $\ker(R)$. By part (1) of this theorem, $\varphi(I)$ is an ideal in R'. We define a map $f : R/I \to R'/\varphi(I)$ by, for any $x \in R$, $f : x + I \mapsto \varphi(x) + \varphi(I)$. We first check that this is well-defined. Suppose $x, y \in R$ are such that $x + I = y + I$. Then $\exists z \in I$ such that $x = y + z$. Thus, $\varphi(x) = \varphi(y) + \varphi(z)$, so since $\varphi(z) \in \varphi(I)$ we get $f(x + I) = f(y + I)$, as required. Next, we check that this is a homomorphism. Suppose that $x, y \in R$. Then $f((x + y) + I) = \varphi(x + y) + \varphi(I) = (\varphi(x) + \varphi(y)) + \varphi(I) = (\varphi(x) + \varphi(I)) + (\varphi(y) + \varphi(I)) = f(x) + f(y)$, and $f(xy + I) = \varphi(xy) + \varphi(I) = \varphi(x)\varphi(y) + \varphi(I) = (\varphi(x) + \varphi(I))(\varphi(y) + \varphi(I)) = f(x)f(y)$, as required. Finally, we show that this is an isomorphism. Pick any $x, y \in R$ and suppose that $f(x + I) = f(y + I)$. Then $\varphi(x), \varphi(y)$ differ only by an element of $\varphi(I)$, and hence $\varphi(x - y) \in \varphi(I)$. Thus, by part (2) of this theorem, $x - y \in I \Rightarrow x + I = y + I$, making f injective. Surjectivity follows from the surjectivity of φ. □

Corollary 3.4.3.1. *Suppose that $I \subset J$ are both ideals in a ring R. Then*

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

Proof. This is just the third part of the proceeding theorem applied to the surjective homomorphism $p : R \to R/I$ (the projection map). □

Theorem 3.4.4 (Third Fundamental Theorem of Homomorphisms)**.** *Suppose S is a subring and I an ideal in a ring R. Then $S + I = \{x + y \mid x \in S, y \in I\}$ is a subring of R containing I, $S \cap I$ is an ideal of S, and*

$$\frac{S}{S \cap I} \cong \frac{S + I}{I}$$

Proof. The proofs of $S + I$ being a subring and $S \cap I$ an ideal of S are left to the reader (the proof is a direct verification). For the last part, we define our map by $f : s + (S \cap I) \mapsto s + I$. That this is a homomorphism if it is well-defined is immediate, so we just check that it's well-defined and bijective. For well-defined, suppose that $s + (S \cap I) = s' + (S \cap I)$. Then $\exists z \in S \cap I$, and in particular $z \in I$, such that $s = s' + z \Rightarrow s + I = s' + I$, as required. For injectivity, suppose that $s + I = s' + I$. Then $\exists z \in I$ such that $s = s' + z$. Since $s - s' = z$, it follows that $z \in S$, so $z \in S \cap I$ and hence $s + (S \cap I) = s' + (S \cap I)$, as required. Finally, we do surjectivity. Pick any $x + y \in S + I$. Then $(x + y) + I = x + I = f(x + (S \cap I))$, making f surjective. □

It's worth at the end here taking a moment to compare these theorems to those in section 2.4, and noting any similarities or differences between them. In fact, there's a sense in which the two sets of theorems are in fact identical, which will be explored further in ADD REFERENCE.

3.5 Field of Fractions

Note. **For the rest of the chapter, all rings are assumed to be commutative unless otherwise stated.**

This is following a similar section in [Jac09], although it has been re-written significantly. The question we explore here is quite simple. Given an arbitrary domain, can we embed it into a field? The answer turns out to be no in general, but it turns out that for commutative rings we can always do this. The natural construction to prove this is actually much more intuitive than one may think. Indeed, at this point you've probably seen a construction of the rational numbers from the integers. This will, in fact, work for any ring.

Definition 3.5.1. The field of fractions of a ring R, denoted $FF(R)$, is the set

$$\{(a, b) \in R \times R^* \mid b \neq 0\}/\sim$$

where $\sim$ is the equivalence on $R \times R^*$ given by $(a, b) \sim (c, d) \iff ad = bc$, equipped with binary operations

$$[(a, b)] + [(c, d)] = [(ad + bc, bd)] \qquad [(a, b)] \cdot [(c, d)] = [(ac, bd)]$$

Proposition 3.5.2. *$FF(R)$ is a well-defined field, with zero element $[(0, 1)]$ and identity $[(1, 1)]$ which R embeds into via the map $x \mapsto (x, 1)$.*

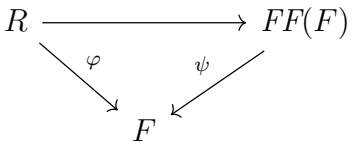
Note. We often refer to the map in the above proposition as the natural embedding of a domain into its field of fractions (although this is technically bad form and shouldn't be done). Since $\mathrm{FF}(R)$ is a field, we will also (suggestively) denote $[(a,b)]$ by a/b or $\frac{a}{b}$. In this notation, the above operations become

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

Proof. Left as an exercise to the reader (it's good practice for working with fields of fractions in calculations, but not particularly enlightening from a conceptual standpoint). The important thing to note is that $(a/b)^{-1} = b/a$. □

Of course, domains can be embedded into multiple fields. What makes the field of fractions special is that it is, in a sense, the smallest such field. This is characterized by the following *universal property*[3].

Theorem 3.5.3. *Suppose R is a domain embedded in some field F via $\varphi : R \hookrightarrow F$. Then there exists a unique $\psi \in Hom(FF(R), F)$ such that the following diagram commutes*



where the unlabelled arrow is he the natural embedding.

Proof. To start, suppose such a $\psi \in \mathrm{Hom}(\mathrm{FF}(R), F)$ existed. Then we'd require that $\varphi(a) = \psi(a/1)$. Furthermore, it would follow that

$$\psi(a/b) = \psi\Big(\frac{a}{1}\frac{1}{b}\Big) = \psi(a/1)\psi(b/1)^{-1} = \varphi(a)\varphi(b)^{-1}$$

But this completely characterizes ψ, making it the desired unique homomorphism if it is well-defined. So, we just need to check that it is in fact well-defined. To that end, we first note that for $b \neq 0$, $\varphi(b) \neq 0$ (as φ is injective), and hence $\varphi(b)^{-1}$ is well-defined. Thus, $\varphi(a)\varphi(b)^{-1}$ is well-defined. To end off then, we just need to check that $a/b = c/d \Rightarrow \psi(a/b) = \psi(c/d)$. But since $ad = bc$, we get

$$\varphi(ad) = \varphi(bc) \Rightarrow \varphi(a)\varphi(b)^{-1} = \varphi(c)\varphi(d)^{-1}$$

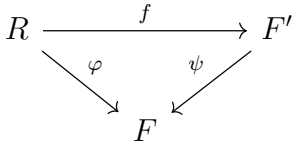
as required. □

In fact, Theorem 3.5.3 implies something a little stronger, that $\mathrm{FF}(R)$ is characterized up to *unique isomorphism*[4]. By this, we mean the following.

[3]These are very important, but we won't worry too much about them at the moment. They will be formally introduced in chapter 7.

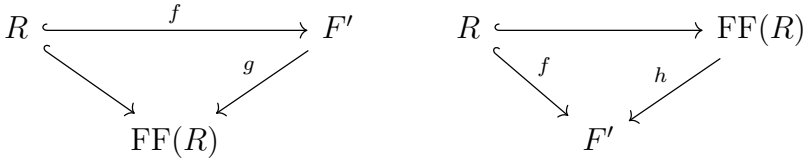
[4]This is the actual universal property.

Corollary 3.5.3.1. *Suppose F' is a field, and $f : R \hookrightarrow F'$ an embedding of a domain. If, for every field F and homomorphism $\varphi : R \to F$ there exists a unique $\psi \in Hom(F', F)$ such that the following diagram commutes*

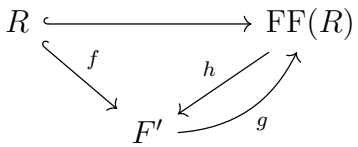


Then there exists one, and only one, isomorphism $g : FF(R) \to F'$ such that $\varphi = g \circ \iota$ always holds.

Proof. If F' satisfies the above property, then there exist by Theorem 3.5.3 unique homomorphisms g, h such that the following diagrams commute



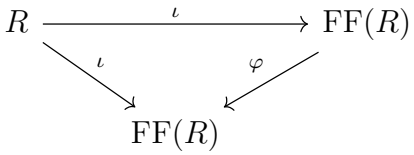
where the unlabelled arrows are the natural embedding. Combining these diagrams, we get one larger commutative diagram



Naming the natural inclusion ι, we see that this implies

$$\iota = g \circ f, h \circ \iota = f \Rightarrow \iota = g \circ h \circ \iota$$

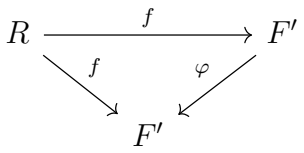
Since ι is injective, we can conclude that $(g \circ h)|_{\mathrm{Im}(\iota)} = \mathrm{Id}_{\mathrm{Im}(\iota)}$. But of course by Theorem 3.5.3 there exists a unique $\varphi \in \mathrm{Hom}(\mathrm{FF}(R), \mathrm{FF}(R))$ such that the following diagram commutes



Since $\varphi = \mathrm{Id}_{\mathrm{FF}(R)}$ works, it follows that $\varphi = \mathrm{Id}_{\mathrm{FF}(R)}$. But of course and φ will do as long as $\varphi|_{\mathrm{Im}(\iota)} = \mathrm{Id}_{\mathrm{Im}(\iota)}$, so it follows that $(g \circ h)|_{\mathrm{Im}(\iota)} = \mathrm{Id}_{\mathrm{Im}(\iota)} \Rightarrow g \circ h = \mathrm{Id}_{\mathrm{FF}(R)}$. Similarly, we also see that

$$f = h \circ g \circ f$$

Since f is injective, we can conclude that $(h \circ g)|_{\mathrm{Im}(f)} = \mathrm{Id}_{\mathrm{Im}(f)}$. But of course by assumption there exists a unique $\varphi \in \mathrm{Hom}(F', F')$ such that the following diagram commutes



Since $\varphi = \mathrm{Id}_{F'}$ works, it follows that $\varphi = \mathrm{Id}_{\mathrm{FF}(R)}$. But of course and φ will do as long as $\varphi|_{\mathrm{Im}(f)} = \mathrm{Id}_{\mathrm{Im}(f)}$, so it follows that $(h \circ g)|_{\mathrm{Im}(f)} = \mathrm{Id}_{\mathrm{Im}(f)} \Rightarrow h \circ g = \mathrm{Id}_{F'}$. Thus, h is the desired isomorphism. That it is unique (in the sense of the corollary statement) follows immediately from the uniqueness of our original choice of h. □

Don't feel too worried if that proof seemed overwhelming or hard to follow, it's our first instance of a technique known as *diagram chasing* which has a habit of being hard formally write out. Take as much time as you need to understand the above proof before moving on.

As one final note, we talked a lot about how to turn rings into fields, but what about recognizing which rings are already fields? To do this, we have the following useful result.

Proposition 3.5.4. *A ring R is a field if and only if its only two ideals are $(0) = \{0\}$ and $(1) = R$.*

Proof. First, suppose that R is a field, and $I \subset R$ an ideal. If $I \neq (0)$, then we can find some non-zero $x \in I$. x is a unit, so for any other $y \in R$ we get $(yx^{-1})x = y \in I$, and hence $I = R$. Now, suppose that there exists some non-zero $x \in R$ which is not a unit. Since R is commutative, $(x) = \{rx \mid r \in R\}$. Thus, since x is not a unit, $rx \neq 1$ always and hence $1 \notin (x)$, so $(x) \neq (1)$. Since $x \neq 0$, $(x) \neq (0)$. □

3.6 Factorial Monoids

Most of the remaining (required) sections of this chapter are all on polynomial rings. Our main concern for polynomial rings, in general, is similar to that of polynomial functions. Namely, we wish to find roots of the polynomials. However, there is an issue with this. Namely, when we view polynomials as rings instead of functions, evaluation becomes a bit of a trickier topic[5]. As such, we wish to find roots without considering evaluation, which leads naturally to the idea of factoring polynomials. To that end, we will take a step back now to understand factoring in a more general context, following a similar section in [Jac09].

For this section, we will work exclusively with commutative monoids M satisfying the cancellation law, that is in the monoid $xy = xz \Rightarrow y = z$.

Definition 3.6.1. For $a, b \in M$, we say that $a \mid b$ (a divides b or is a factor of b) if there exists some $z \in M$ such that $b = za$. a is a proper factor of b if $a \mid b$ but $b \nmid a$. An element $b \in M$ is irreducible if its only proper factors are units, and is prime if $b \mid cd$ implies that $b \mid c$ or $b \mid d$, for any $c, d \in M$.

[5]It can be done without too much difficulty, but you have to be careful.

Note. If $a \mid b$ and $b \mid a$, then one can use the cancellation law to show that $a = zb$, where $z \in M$ is a unit. Also, prime elements are necessarily irreducible. Indeed, suppose that $p \in M$ is prime and $a \mid p$, with $ab = p$. Then $p \mid a$ or $p \mid b$. In the first case we've shown that a is not a proper factor, and we're done. In the second, cancellation law implies that a is a unit, and hence we're done. Note, however, that irreducible elements need not be prime in general.

Definition 3.6.2. A factorization of an element $a \in M$ is an expression of the form $a = p_1 \cdots p_n$, where $p_k \in M$ are irreducible. Such a factorization is called unique if any other factorization can be obtained from the original one by re-ordering elements and multiplying by units.

This finally allows us to define our desired objects.

Definition 3.6.3. M is factorial if every element of M has a unique factorization.

The classic example of one of these is $\mathbb{Z}^*$ under multiplication. We'll now spend the rest of this section defining some equivalent conditions for monoids to be factorial. The first two conditions we'll explore are the following.

Definition 3.6.4. A monnoid M is said to satisfy

1. The ascending chain condition (ACC) if there exists no infinite sequence of elements $a_i \in M$ such that a_{i+1} is a proper factor of a_i.

2. The primeness condition if every irreducible element of M is prime.

Theorem 3.6.5. *A monoid M is factorial if and only if it satisfies the ACC and primeness conditions.*

Proof. First, suppose that M is factorial. Suppose that $x, y, z \in M$ are such that x is irreducible and $x \mid yz$. Let $w \in M$ be the element such that $xw = yz$. By the uniqueness of irreducible decompositions, and since x is an irreducible element, we conclude that x (up to multiplication by some unit) must be in the irreducible decomposition of y or z, and hence $x \mid y$ or $x \mid z$. x is therefore prime, and M satisfies the primeness condition. Now, suppose $\{a_i\}_{i \in \mathbb{N}}$ is a sequence of elements in M violating the ACC. Take an irreducible decomposition of a_1, say $a_1 = p_1 \cdots p_n$, and one of a_2, say $a_2 = q_1 \cdots q_m$, where we assume that p_j, q_j are not units (if they were then a_2 would not be a proper factor of a_1). Let $z \in M$ be such that $a_2 z = a_1$. Then by the uniqueness of irreducible decompositions and primeness condition, $q_1 \mid p_1$ or $q_1 \mid p_2 \cdots p_n$. If $q_1 \mid p_1$, then it is just a unit multiple of p_1 as p_1 is irreducible. Otherwise, we can repeat this argument on $p_2 \cdots p_n$ and thereon, finding some $1 \leq j \leq n$ such that $q_1 = up_j$, where $u \in M$ is a unit. We may then apply cancellation law to remove the factor q_1 from both sides, and repeat the argument with q_2, eventually concluding that for each q_j we can find some p_{k_j} and unit $u_j \in M$ such that $q_j = u_j p_{k_j}$ and each k_j is distinct. Since the number of irreducible factors in the decomposition of an element in a factorial monoid is unique, we may conclude (cancelling the q_j on both sides) that since a_2 is a proper factor of a_1, it must have strictly fewer irreducible factors in its decomposition compared

to a_1. We may then repeat this argument with a_2, a_3 and so on, eventually finding some $r \in \mathbb{N}$ such that a_r is irreducible. But then a_r cannot have any proper factors, violating our assumption about the nature of $\{a_i\}_{i\in\mathbb{N}}$. Therefore, M satisfies the ACC.

Now, suppose that M satisfies the ACC and primeness conditions. We first show that any $x \in M$ has an irreducible decomposition. If x is irreducible than this is immediate. Otherwise, pick some non-unit irreducible factor $a_1 \in M$ and element $b_1 \in M$ such that $x = a_1b_1$. Then b_1 is a proper factor of x. Indeed, suppose that $x \mid b_1$. Let $y \in M$ be such that $xy = b_1$. Then $a_1b_1y = b_1 \Rightarrow a_1y = 1$ and hence a_1 is a unit contrary to our assumption. If b_1 is irreducible, then a_1b_1 is an irreducible decomposition, and we're done. Otherwise, we repeat the above process on b_1, and continue until all factors are irreducible. This process must terminate, as otherwise the b_i sequence constructed along the way would violate the ACC. Thus, every element of M has an irreducible decomposition. We finish by proving the uniqueness of these decompositions. Suppose $x = p_1 \cdots p_n = q_1 \cdots q_m$ are two irreducible decompositions. Since $p_1 \mid q_1 \cdots q_m$, we conclude by the primeness condition that $p_1 \mid q_1$ or $p_1 \mid q_2 \cdots q_m$. We continue this process, repeating the same argument as in the previous part of the proof, to pair up each p_i with a distinct q_j it is a unit multiple of. Cancelling all the p_i must then leave us with only units, giving the uniqueness of the decomposition and making M a factorial domain. □

There's one more idea we'd like to generalize before we move on, namely the concept of greatest common divisors and least common multiples.

Definition 3.6.6. For any two elements $x, y \in M$, we call $z \in M$

1. A least common multiple (LCM) of x, y (or an element of $\operatorname{lcm}(x, y)$) if $x, y \mid z$ and, for any $w \in M$, $x, y \mid w \Rightarrow z \mid w$.

2. A greatest common divisor (GCD) of x, y (or an element of $\gcd(x, y)$) if $z \mid x, y$ and, for any $w \in M$, $w \mid x, y \Rightarrow w \mid z$.

Note. One can check that, in $\mathbb{Z}$, these are equivalent to our traditional notions of LCM and GCD. Furthermore, LCM and GCD of a pair of elements are unique up to units.

Theorem 3.6.7. *Any pair of elements x, y in a factorial monoid M have a GCD and LCM.*

Proof. First, suppose that x is a unit. Then any element of M dividing x is also a unit, $x \mid y$, and one can quickly check that $x \in \gcd(x, y), y \in \operatorname{lcm}(x, y)$. Now, suppose that x, y are not units. Let $x = p_1 \cdots p_n, y = q_1 \cdots q_m$ be their irreducible decompositions, where we assume that the irreducible factors are not units. Then by a similar argument as in Theorem 3.6.5, any element of M dividing x, y must have as irreducible factors only factors appearing in x, y (up to multiplication by units), and none of these factors can appear more times than they did in x, y. It follows that taking the product of all common irreducible factors of x, y (counting multiplicity and considering factors up to multiplication by units) gives a GCD of x, y. A similar argument shows that taking the product of the minimum number of irreducible factors needed to have all those in x, y gives an LCM. □

Note. This result and the GCD/LCM constructed are just extensions of the same result CITE.

Note. The proof here is more of a sketch, as the author is on summer vacation and couldn't be bothered.

It turns out that the GCD and factoriality of a monoid are closely related. This is related to the following two results, both of which are proven in [Jac09] and will not be proven here.

Proposition 3.6.8. *1. If any pair of elements in a monoid M have a GCD (this is referred to as the GCD condition), then so does any finite collection of elements in M.*

2. The GCD condition implies the primeness condition.

Corollary 3.6.8.1. *M is factorial if and only if it satisfies the ACC and GCD conditions.*

Proof. If M is factorial, then it satisfies ACC by Theorem 3.6.5 and GCD by Theorem 3.6.7. If M satisfies ACC and GCD, then by proposition 3.6.8 it satisfies the primeness condition, and hence by Theorem 3.6.5 is factorial. □

3.7 PIDs and Euclidian Domains

This is again following a similar section in [Jac09]. We start with a basic observation.

Proposition 3.7.1. *Let R be a domain. Then for any $x, y \in R$, $x \mid y$ if and only if $(y) \subset (x)$.*

Proof. If $x \mid y$, then there exists some $z \in R$ such that $xz = y$, and hence $y \in (x) \Rightarrow (y) \subset (x)$. If $(y) \subset (x)$, then $y \in (x)$, so there exists some $z \in R$ such that $xz = y$. Therefore, $x \mid y$. □

Note. By $x \mid y$, we mean here that $x \mid y$ in the commutative monomial $(R, \cdot)$, which since R is a domain satisfies the cancellation law.

Essentially, the above proposition says that we may study ideals generated by one element (these are called *principle ideals*) instead of studying divisibility directly. Ideally, then, we'd like all the ideals in R to be principle, and we call R a *principle ideal domain* (PID) if it satisfies this. Specifically, we'd like to find rings such that $(R, \cdot)$ is factorial. We call such rings *unique factorization domains* (UFDs).

Theorem 3.7.2. *If R is a PID, then R is a UFD.*

Proof. First, we show that R satisfies the ACC. Suppose

$$(a_1) \subset (a_2) \subset \cdots \subset (a_n) \subset \cdots$$

is a chain of principle ideals in R. Then $A = \bigcup_{i \in \mathbb{N}} (a_i)$ is an ideal, and hence since R is a PID there exists some $x \in R$ such that $A = (x)$. Thus, $x \in (a_n)$ for some $n \in \mathbb{N}$, but also by definition $(a_n) \subset (x)$. We conclude that $(a_n) = (x)$, so a_n, x differ only by a unit. The ACC follows from this and proposition 3.7.1. Next, we show that R satisfies the GCD condition. In particular, pick any $x, y \in R$. Then since R is a PID, there exists some $z \in R$ such that

$(x, y) = (z)$. Since $(x), (y) \subset (z)$, $z \mid x, y$. Now, suppose that $w \in R$ is some other element such that $w \mid x, y$. Then $(x), (y) \subset (w)$, so since ideals are closed under addition we conclude that $(x, y) \subset (w) \Rightarrow (z) \subset (w) \Rightarrow w \mid z$. Thus, z is the desired GCD. The result now follows by corollary 3.5.3.1. □

Actually proving that rings are PIDs can be somewhat tricky, but (as we'll see later in this section) having a long division algorithm in R like that of $\mathbb{Z}$ is sufficient to make a ring a PID. Thus, we generalize long division.

Definition 3.7.3. A domain R is Euclidian if there exists a map $\delta : R \to \mathbb{Z}^*$ such that for any $a, b \in R^*$, there exists some $q, r \in R$ such that $a = bq + r$, where $\delta(r) < \delta(b)$.

Note. δ is our way of measuring the "size" of our remainder r. In the case of $\mathbb{Z}$, we'd have $\delta(x) = |x|$.

Theorem 3.7.4. *Euclidian domains are PIDs.*

Proof. Suppose R is Euclidian. Let I be any ideal in R, and suppose that $I \neq (0)$. Let $b \in I$ be a non-zero element in I such that $\delta(b)$ is minimal. Suppose $\exists a \in I$ such that $b \nmid a$. Since R is Euclidian, there exist $q, r \in R$ such that $a = qb + r$, and $\delta(r) < \delta(b)$. But $b \nmid a \Rightarrow r \neq 0$, and $a - qb \in I \Rightarrow r \in I$, so this contradicts the minimality of $\delta(b)$. Thus, every element of I is a multiple of b, so $I = (b)$. □

Corollary 3.7.4.1. *Euclidian domains are UFDs.*

3.8 Polynomial Rings

We follow the results of [Jac09] here again, although the organization of the material has been significantly changed.

Polynomial rings are, without a doubt, the most important example of rings for algebra. In a way, the entirety of **??** is dedicated to the study of polynomial rings and their structure. So without further ado, let's get to it.

Definition 3.8.1. Let R be a ring. The polynomial ring over R in one variable, denoted $R[x]$, is the set $R_c^{\mathbb{Z}_{\geq 0}}$ (infinite sequences in R with finitely many non-zero elements), with element-wise addition and multiplication defined by

$$((a_i)_{i \in \mathbb{Z}_{\geq 0}} (b_i)_{i \in \mathbb{Z}_{\geq 0}})_j = \sum_{i+k=j} a_i b_k$$

Note. Our zero here is $(0, 0, 0, \dots)$, and our identity is $(1, 0, 0, \dots)$.

It's not immediate from this definition the connection between this ring and polynomials as we know them. To make this connection more clear, we usually adopt the following notation. First, denote the sequence with a 1 in the nth position by x^{n-1} for $n \geq 1$. One can note that

$$(c, 0, 0, \dots) \cdot (a_0, a_1, \dots) = (ca_0, ca_1, \dots)$$

Thus, any sequence can be written uniquely[6] in the following form

$$(a_0, a_1, \dots) = (a_0, 0, \dots) + (a_1, 0, \dots)x + (a_2, 0, \dots)x^2 + \cdots$$

Denoting $(c, 0, \dots)$ by just c, this becomes

$$(a_0, a_1, \dots) = a_0 + a_1x + a_2x^2 + \cdots$$

making the connection much more clear. In fact, we will always choose to use this notation, as it makes everything much more intuitive. One can check that multiplying out two expressions of the above form in the way that you normally would for polynomials will give the correct result, only furthering this connection.

Note. This polynomial ring is actually distinct from the ring of polynomial functions over R, which we'll cover a bit later in this section.

We can then continue to generalize this to multivariable polynomial rings.

Definition 3.8.2. We define the multivariable polynomial ring over R in the following inductive manner. That is, we denote the polynomial ring in n variables over R by $R[x_1, \dots, x_n]$, and define it by $R[x_1, \dots, x_n] = R[x_1, \dots, x_{n-1}][x_n]$.

It's at this point where the notation we've been using, writing polynomial rings in the same way as polynomial functions, becomes incredibly convenient. For example, we'd get that in $R[x_1, x_2]$

$$\begin{aligned}((a, b, c, 0, \dots), (a, b, 0, \dots), (a, 0, \dots), (0, \dots), \dots) &= (a + bx_1 + cx_1^2) + (a + bx_1)x_2 + ax_2^2 \\ &= a + bx_1 + cx_1^2 + ax_2 + bx_1x_2 + ax_2^2\end{aligned}$$

We call terms of the form $x_1^{k_1} \cdots x_n^{k_n}$ *monomials* in $R[x_1, \dots, x_n]$. Like in $R[x]$, any element of $R[x_1, \dots, x_n]$ can be written uniquely as the sum of finitely many monomial elements summed together, with each monomial multiplied by some non-zero coefficient in R[Jac09][7].

We move away from polynomials now, for a moment, and talk instead about the related concept of *adjoined rings*.

Definition 3.8.3. Let R be a subring of a ring S, and let $U \subset S$. Then R adjoin U, denoted $R[U]$, is the subring of S generated by $R \cup U$.

Proposition 3.8.4. *Suppose R is a subring of S, and $U, V \subset S$. Then $R[U][V] = R[U \cup V]$.*

Proof. Since $R[U \cup V]$ is a subring of S containing R and U, $R[U] \subset R[U \cup V]$. Thus, $R[U], V$ are contained in $R[U \cup V]$, so $R[U][V] \subset R[U \cup V]$. Furthermore, $R[U][V]$ contains R, U, and V, so $R[U \cup V] \subset R[U][V]$. □

[6] This technically needs to be proven, but I don't think the proof is particularly enlightening or complicated.

[7] The proof of this is an inductive argument, and rather tedious.

Proposition 3.8.5. *Suppose R is a subring of S, and $u \in S$. Then $R[u] = R[\{u\}]$ is the subring of S composed of expressions of the form*

$$\sum_k a_k u^k, a_k \in R, \textit{finite sums}$$

Proof. That all expressions of this form are in $R[u]$ is immediate. Thus, since (as can be quickly checked) since this is a subring containing R and u, we get the desired result. □

Note. The above result suggests a strong connection between $R[x]$ and $R[u]$, the latter is a way of evaluating the polynomial expressions in the former. This is also why we take the blatant abuse of notation of denoting them in the same way.

Note. The above proposition in fact works for any $R[U]$, extending the expressions in the obvious way. The proof is identical.

The above observation will now be expanded upon in the following extremely important theorem. For this theorem, we take the convention of the "constants" in our standard notation for $R[x_1, \ldots, x_n]$ being an embedding of R in $R[x_1, \ldots, x_n]$.

Theorem 3.8.6. *Let R be a ring, and S any ring containing R. Then for any $n \in \mathbb{N}$ and $u_1, \ldots, u_n \in S$, there exists a unique homomorphism $\varphi : R[x_1, \ldots, x_n] \to S$ fixing R and taking $x_i \mapsto u_i$.*

Proof. Since $R[x_1, \ldots, x_n] = R[x_1, \ldots, x_{n-1}][x_n]$, it suffices by induction to prove this for the case $n = 1$. Note that any $a \in R[x]$ can be written uniquely in the form

$$a = \sum_{k \geq 0} a_k x^k$$

where $a_k \in R$ and only finitely many $a_k \neq 0$. We'll define $\varphi : R[x] \to S$ by

$$\varphi(a) = \sum_{k \geq 0} a_k u^k$$

It is fairly simple to check that this is in fact a ring homomorphism. For uniqueness, note that if $\varphi|_R = \mathrm{Id}_R$ and $\varphi(x) = u$, then since φ is a ring homomorphism

$$\varphi(a) = \varphi(a) = \sum_{k \geq 0} \varphi(a_k)\varphi(x)^k = \sum_{k \geq 0} a_k u^k$$

□

Note. What we mean by "R is a subring of S" can often be a bit loose. We make no distinction between R being a subring of S, or R embedding into S via a ring homomorphism. Indeed, identifying R with its embedding, we can see that there really is no difference between the two situations.

Corollary 3.8.6.1. *Fix $n \in \mathbb{N}$. Suppose K is any ring containing R and distinguished elements $y_1, \ldots, y_n \in K$ such that*

1. $R[y_1, \ldots, y_n] = K$
2. *For any other ring S containing R and $u_1, \ldots, u_n \in S$ there exists a unique homomorphism $\varphi : K \to S$ which fixes R and satisfies $\varphi(y_i) = u_i$.*

Then $K \cong R[x_1, \ldots, x_n]$.

Proof. By assumption, there exists a unique $\varphi \in \mathrm{Hom}(K, R[x_1, \ldots, x_n])$ fixing R such that $\varphi(y_i) = x_i$. Furthermore, there exists by Theorem 3.8.6 a unique $\psi \in \mathrm{Hom}(R[x_1, \ldots, x_n], K)$ fixing R such that $\psi(x_i) = y_i$. Then $(\varphi \circ \psi)(x_i) = x_i$ and $\varphi \circ \psi$ fixes R, so by the uniqueness property in Theorem 3.8.6 $\varphi \circ \psi = \mathrm{Id}_{R[x_1,\ldots,x_n]}$. An identical argument shows that $\psi \circ \varphi = \mathrm{Id}_K$, so φ is an isomorphism. □

Note. What the above argument is really saying is that, up to isomorphism, there's only one way to define a multivariable polynomial ring over R.

Corollary 3.8.6.2. *For any permutation $\sigma \in S_n$, there exists a unique automorphism $\varphi_\sigma \in \mathrm{Isom}(R[x_1, \ldots, x_n])$ such that $\varphi(x_i) = x_{\sigma(i)}$.*

Proof. Left to the reader. This has essentially the same proof as corollary 3.8.6.1. □

We call the homomorphism given by Theorem 3.8.6 the *evaluation homomorphism*. Our main goal from now on will be to understand the kernels of these homomorphisms. Indeed, we have a good understanding of the structure of $R[x]$, and $R[u] \cong R[x]/\ker\varphi$, so understanding the kernel of evaluation homomorphisms teaches us a lot about the element of rings containing R. It will also, unsurprisingly, have deep connections to identification of roots of polynomials. On that note, let's actually define these polynomial functions.

Definition 3.8.7. The ring of polynomial functions in n variables over R, denoted $\mathcal{P}_n(R)$, is the subset of the ring of functions from $R^n \to R$ of the form

$$f(u_1, \ldots, u_n) = \sum_{(k_1,\ldots,k_n)} a_{k_1,\ldots,k_n} u_1^{k_1} \cdots u_n^{k_n}$$

where $a_{k_1,\ldots,k_n} \in R$ and the above sum is finite.

In order to understand the connections between these, our polynomial rings, and roots of polynomials, we'll need to study the factoring of polynomials.

3.9 Factoring Polynomials

Again, this follows some similar sections in [Jac09]. In order to begin factoring, we first need to understand the notion of degree.

Definition 3.9.1. Let $f \in R[x]$ be a polynomial. Then we may write, for some $n \in \mathbb{N}$ and $a_n \neq 0$

$$f = \sum_{k=0}^{n} a_k x^k$$

We define the degree of f, denoted $\deg(f)$, to be n, and call a_n the leading coefficient of f. By convention, we define that $\deg(0) = -\infty$, where $-\infty$ has the expected arithmetic properties.

Note. This is well-defined by the uniqueness of this method of expressing f. For multivariable polynomials, the notion of degree gets a little tricker. One can either look at their degree in a particular variable, or their total degree. We'll look at that a bit more in the next section.

A couple of the properties of degree are fairly immediate, and will not be proven here.

Proposition 3.9.2. *Suppose $f, g \in R[x]$. Then*

1. $\deg(f+g) \leq \max(\deg(f), \deg(g))$
2. $\deg(fg) = \deg(f)\deg(g)$

Using the notion of degree, we can immediately start proving some useful results.

Proposition 3.9.3. *If R is a domain, then $R[x_1, \ldots, x_n]$ is a domain.*

Proof. Since $R[x_1, \ldots, x_{n-1}][x_n] = R[x_1, \ldots, x_n]$, it suffices to show that $R[x]$ is a domain. Indeed, suppose that $f, g \in R[x]^*$. Then $\deg(f), \deg(g) \geq 0$, so $\deg(fg) \geq 0 \Rightarrow fg \neq 0$. □

The next result is perhaps one of the most fundamental in this section, namely that polynomial long division can be extended to arbitrary polynomial rings.

Theorem 3.9.4. *Suppose $f, g \in R[x]$, with $g \neq 0$. Let $m = \deg(g)$ and $b_m \neq 0$ be the leading coefficient of g. Then there exists some $k \in \mathbb{N}$, $q, r \in R[x]$ with $\deg(r) < \deg(g)$ such that*

$$b_m^k f = qg + r$$

Proof. If $\deg(f) < \deg(g)$, then we simply take $q = 0, r = b_m f, k = 1$ to get the desired result. Otherwise, define

$$f_1 = b_m f - a_n x^{n-m} g$$

where $n = \deg(f)$, and a_n is the leading coefficient of f. It's clear that $\deg(f_1) \leq \deg(f) - 1$. If $\deg(f_1) < \deg(g)$, then we're done. Otherwise, we can repeat this process with f_1, getting an $f_2, \ldots, f_\ell \in R[x]$, continuing until $\deg(f_\ell) < \deg(g)$. This is guaranteed to terminate, since $\deg f_[k+1) \leq \deg(f_k) - 1$. In the end, we get

$$f_\ell = b_m f_\ell - a^{(\ell-1)} x^{\deg(f_{\ell-1} - m)} g$$

where $a^{(k)}$ is the leading coefficient of f_k. But in turn we know that

$$f_{\ell-1} = b_m f_{\ell-2} - a^{(\ell-2)} x^{\deg(f_{\ell-2} - m)} g$$

and so on. Expanding out, this gives

$$f_\ell = b_m^2 f_{\ell-2} - (a^{(\ell-1)} x^{\deg(f_{\ell-1} - m)} + a^{(\ell-2)} x^{\deg(f_{\ell-2} - m)}) g$$

Continuing this process, and calling the collected terms which multiply g by $q \in R[x]$, we see that

$$f_\ell = b_m^\ell f - qg \Rightarrow b_m^\ell f = qg + f_\ell$$

Since $\deg(f_\ell) < \deg(g)$, this completes the proof. □

Note. This proof also gives you an algorithm for calculating this "long division".

If b_m is a unit, then since $b_m^k \neq 0$ we get the following, more familiar result.

Corollary 3.9.4.1. *If $f, g \in R[x]$, and $g \neq 0$, then there exists unique $q, r \in R[x]$ with $\deg(r) < \deg(g)$ such that*

$$f = qg + r$$

Furthermore, in FF($R[x]$), we get

$$\frac{f}{g} = q + \frac{r}{g}$$

Proof. Existence is given by Theorem 3.9.4 and dividing out by the unit. FOr uniqueness, suppose q_1, r_1 and q_2, r_2 were two such pairs. Then

$$q_1 g + r_1 = q_2 g + r_2 \Rightarrow (q_1 - q_2)g = r_2 - r_1$$

Taking the degree of both sides, we get

$$\deg(q_1 - q_2)\deg(g) \leq \max(\deg(r_2), \deg(r_1)) < \deg(g)$$

We are therefore left with two possibilities. First, suppose that $\deg(g) = 0$. Then $\deg(r_1)$, $\deg(r_2) < 0 \Rightarrow r_1 = r_2 = 0$, so $(q_1 - q_2)g = 0 \Rightarrow q_1 = q_2$. Otherwise, we must conclude that $q_1 = q_2$, which in turn implies that $r_1 = r_2$. □

Note. In this case, we call q and r the quotient and remainder of f/g.

Using this, we can start factoring our polynomials properly. In order to do so, we'll need a bit of notation. Suppose $R \subset S$ is a subring, $f \in R[x]$, and $a \in S$. Then we'll use $f(x)$ to denote the polynomial in $R[x]$, and $f(a)$ to denote the *evaluation* (i.e. image under the evaluation homomorphism) of f at a.

Corollary 3.9.4.2 (Remainder Theorem)**.** *Suppose $f(x) \in R[x]$ and $a \in R$. Then there exists a unique $q(x) \in R[x]$ such that*

$$f(x) = (x - a)q(x) + f(a)$$

Proof. By corollary 3.9.4.1, there exist some unique $q(x), r(x) \in R[x]$ such that $\deg(r(x)) < \deg(x - a)$ and

$$f(x) = (x - a)q(x) + r(x)$$

In particular, since $\deg(r(x)) < \deg(x - a) = 1$, $r(x) \in R$ (or more properly its embedding into $R[x]$). Thus, $r(x)$ is fixed by any evaluation homomorphism. In particular, we can then evaluate both sides of the above equation at a to get

$$f(a) = (a - a)q(a) + r(x) \Rightarrow r(x) = f(a)$$

□

We also get the following result immediately from the above corollary.

Corollary 3.9.4.3 (Factor Theorem)**.** *Suppose $f(x) \in R[x]$ and $a \in R$. Then $(x-a) \mid f(x)$ if and only if $f(a) = 0$.*

There are two more results we can get out of these theorems, namely on the number of roots polynomials over fields have and on the structure of polynomial rings over a field.

Definition 3.9.5. Let F be a field, $f(x) \in F[x]$ be such that $\deg(f) > 0$. We call $a \in F$ a root of f if $f(a) = 0$.

Corollary 3.9.5.1. *Let $f(x) \in F[x]$ be a polynomial of degree $n \geq 1$. Then $f(x)$ has at most n distinct roots in F.*

Proof. Let $a_1, \ldots, a_m \in F$ be distinct roots of F. We show, by induction on r, that $\prod_{k=1}^{r}(x - a_r) \mid f(x)$, from which the result immediately follows. The case of $r = 1$ is given by the factor theorem. Now, suppose this holds for some $r \geq 1$ such that $r < m$. Then there exists some $g(x) \in F[x]$ such that

$$g(x)\prod_{k=1}^{r}(x - a_k) = f(x)$$

evaluating both sides at a_{r+1}, we get that since $f(a_{r+1}) = 0$ and $(a_{r+1} - a_k) \neq 0$ for $1 \leq k \leq r$, $g(a_{r+1}) = 0$. Thus, by the factor theorem, there exists some $h(x) \in F[x]$ such that $g(x) = (x - a_{r+1})h(x)$, and so

$$h(x)\prod_{k=1}^{r+1}(x - a_k) = f(x) \Rightarrow \prod_{k=1}^{r+1}(x - a_k) \mid f(x)$$

□

Corollary 3.9.5.2. *If F is a field, then $F[x]$ is a PID.*

Proof. Let $I \subset F[x]$ be an ideal. Let $f(x) \in I$ be a non-zero polynomial of minimal degree. Take any other polynomial $g(x) \in F[x]$. By corollary 3.9.4.1, there exist $q(x), r(x) \in F[x]$ such that $\deg(r) < \deg(f)$ and

$$g(x) = q(x)f(x) + r(x) \Rightarrow r(x) = g(x) - q(x)f(x)$$

Since $r(x) \in I$, we conclude by the minimality of the degree of $f(x)$ that $r(x) = 0$. Thus, $f(x) \mid g(x)$. Since every element of I is divisible by $f(x)$, and $f(x) \in I$, it follows that $I = (f)$. □

Note. This result is false for multivariable polynomial rings, a decent example of this can be found in [Jac09].

There's a strong connection between evaluation homomorphisms and the irreducibility of polynomials we're building up to here, but first we'll need the following definitions.

Definition 3.9.6. Let $R \subset S$ be a subring, and pick $a \in S$. We call a algebraic over R if there exists a monic (i.e. polynomial with leading coefficient 1) $f(x) \in R[x]$ such that $f(a) = 0$. Otherwise, we call it transcendental over R. For algebraic elements, we call a monic polynomial $f(x) \in R[x]$ such that $f(a) = 0$ a minimal polynomial of a over R.

Proposition 3.9.7. *Let $F \subset K$ be a subfield, and pick any $a \in K$ algebraic over F. Then a has a unique minimal polynomial.*

Proof. Note that the set of all polynomials of which a is a root is an ideal I. By corollary 3.9.5.2, $F[x]$ is a PID. Thus, there exists some non-zero $f(x) \in F[x]$ such that $I = (f(x))$. In particular, since we're operating over a field, we can choose for $f(x)$ to be monic. Since any polynomial in I is a multiple of $f(x)$, there is no other monic polynomial in I of degree less than or equal to $f(x)$. □

In this case, we take to calling $f(x)$ the *minimal polynomial* of a over F.

Theorem 3.9.8. *Suppose $F \subset K$ is a subfield, and $u \in K$ is algebraic over F with minimal polynomial $f(x) \in F[x]$. Then $F[u]$ is a field if $f(x)$ is irreducible, and is not a domain otherwise.*

Proof. First, suppose that $f(x)$ is irreducible. Let $\varphi \in \mathrm{Hom}(F[x], K)$ be the evaluation homomorphism. We can first note that $F[u] \cong \varphi(F[x])$, and hence

$$F[u] \cong F[x]/\ker(\varphi)$$

Every ideal in $F[u]$ is therefore the image of an ideal in $F[x]$ containing $\ker(\varphi)$ under the quotient map. By the proof of proposition 3.9.7, $\ker(\varphi) = (f(x))$. Therefore, ideals in $F[u]$ correspond to ideals in $F[x]$ containing $f(x)$. Suppose J were such an ideal. Then since $F[x]$ is a PID and F a field, there exists a monic $g(x) \in F[x]$ such that $J = (g(x))$. Thus, since $f(x) \in (g(x)), g(x) \mid f(x)$. But this implies that $g(x)$ is a unit, and hence $g(x) \in F \Rightarrow g(x) = 1$. Thus, the only two ideals in $F[u]$ are the zero ideal and $F[u]$, making $F[u]$ a field. Now, suppose that $f(x)$ is reducible, say with factoring $f(x) = g(x)h(x)$, where $\deg(g), \deg(h) \geq 1$. By the minimality of the degree of f, $g(u), h(u) \neq 0$. However, $f(u) = g(u)h(u) = 0$. Thus, $F[u]$ is not a domain. □

You may think we're done with factoring, but you'd be wrong. We can, in fact, build up to one last much stronger result. Namely, that if R is a UFD, then so is $R[x]$. To do this, we'll need to introduce the concept of the *content* of a polynomial.

Definition 3.9.9. Let R be a UFD, and $f(x) \in R[x]^*$. Writing

$$f(x) = a_n x^n + \cdots + a_1 x + a_0$$

We define the content of f, denoted $c(f)$, by

$$c(f) = \mathrm{GCD}(a_1, \ldots, a_n)$$

If $c(f)$ is a unit, we call f primitive.

Note. The content is only-well defined up to multiplication by units.

In the following, we will need an identity on GCDs which we will not prove. A proof can be found in [Jac09].

Proposition 3.9.10. *Suppose M is a factorial monoid, and $a, b, c \in M$. Then up to multiplication by units*

1. *$GCD(GCD(a,b),c) = GCD(a,b,c)$*
2. *$GCD(ac,bc) = cGCD(a,b)$*

Note. The first part of this proposition is a little vague, but essentially means that this equality works (up to multiplication by units in M) for any choice of GCD at any point in evaluating the equations.

Using this, we get the following.

Proposition 3.9.11. *Suppose R is a UFD, and $f(x) \in R[x]^*$. Then there exists a primitive polynomial $g(x) \in R[x]^*$ and constant $a \in R$ such that $f(x) = ag(x)$. Furthermore, if $f(x) = bh(x)$ is another such decomposition, then there exists a unit $u \in R$ such that $b = ua$.*

Proof. Take some $a \in c(f)$ (i.e. choose a particular content). Let $f(x) = c_n x^n + \cdots + c_0$. Then by definition, $a \mid c_k$ for each $0 \leq k \leq n$, and setting $d_k = c_k/a$ and $g(x) = d_n x^n + \cdots + d_0$ we get by proposition 3.9.10 that $c(g)$ is a unit, and hence g is primitive. $f(x) = ag(x)$ is therefore the desired decomposition. Now, suppose that $f(x) = bh(x)$ is another such decomposition. Write $h(x) = k_n x^n + \cdot + k_0$. Since h is primitive, we get that since $c(f) = c(ag(x))$, $b \in c(f)$. Thus, there exists $u \in R$ such that $b = ua$. □

We next generalize the content of polynomials to polynomials over fields in the following manner.

Lemma 3.9.12. *Suppose R is a UFD, $F = FF(R)$, and $f(x) \in F[x]^*$. Then there exists some $a \in F$ and primitive polynomial $g(x) \in R[x]^*$ such that $f(x) = ag(x)$. Furthermore, if $f(x) = bh(x)$ is another such decomposition, then there exists a unit $u \in R$ such that $a = bu$.*

Proof. Let $f(x) = c_n x^n + \cdots + c_0$. Since $F = \mathrm{FF}(R)$, there exists some $\alpha \in R^*$ such that $\alpha c_k \in R$ for every $0 \leq k \leq n$. Then $\alpha f(x) \in R[x]^*$, and hence by proposition 3.9.11 there exists some $a \in R$ (in particular $a \in c(\alpha f(x))$) and primitive $g(x) \in R[x]^*$ such that $\alpha f(x) = ag(x)$. Thus, $f(x) = \frac{a}{\alpha} g(x)$ is the desired decomposition. Now, suppose that $ag(x), bh(x)$ are two such decompositions. Then $(\alpha a)g(x) = (\alpha b)h(x)$ are in $R[x]^*$, so by proposition 3.9.11 there exists some unit $u \in R$ such that $\alpha a = u\alpha b \Rightarrow a = ub$. □

In the case of the above lemma, we call the a the *field content* of $f(x)$.

Lemma 3.9.13 (Gauss's Lemma)**.** *The product of primitive polynomials is primitive.*

Proof. Suppose $g(x), h(x)$ are primitive, but $f(x) = g(x)h(x)$ is not. Then there exists an irreducible, and hence prime, $p \in R^*$ such that $p \nmid g(x), h(x)$, but $p \mid f(x)$. Note that since p is prime, $R' = R/(p)$ is a domain[8]. Projecting all out polynomials into $R'[x]$, we get $g(x), h(x) \neq 0$ but $f(x) = 0$. Thus, $R'[x]$ is not a domain, which contradicts proposition 3.9.3. □

[8] We technically have not proven this yet, but it is not too hard to check

We can now, finally, start proving the relevant results.

Theorem 3.9.14. *If $f(x) \in R[x]$ has degree at least one, where R is a UFD and $F = FF(D)$, then $f(x)$ is irreducible in $R[x]$ if and only if it is irreducible in $F[x]$.*

Proof. If $f(x)$ is irreducible in $F[x]$, then irreducibility in $R[x]$ is immediate. Now, suppose that $f(x)$ is irreducible in $R[x]$, and is of degree at least one. By lemma 3.9.12, there exists some $a \in F$ and primitive $g(x) \in R[x]$ such that $f(x) = ag(x)$. Suppose f were reducible in $F[x]$, say $f(x) = h(x)k(x)$. We note that since all non-zero constants in F are invertible, $\deg(h), \deg(k) \geq 1$. By lemma 3.9.12, there exist some $b, c \in F$ and primitive $f_1(x), f_2(x) \in R[x]$ such that $h(x) = bf_1(x), k(x) = cf_2(x)$. By Gauss's lemma, $f_1(x)f_2(x)$ is primitive, so since $f(x) = (bc)(f_1(x)f_2(x))$ we conclude that there exists some unit $u \in R$ such that $uf_1(x)f_2(x) = f(x)$. But then $f(x)$ would be reducible in $R[x]$, contradicting our assumption. □

Theorem 3.9.15. *If R is a UFD, then so is $R[x]$.*

Proof. Suppose $f(x) \in R[x]^*$ is irreducible. Then it is irreducible in $F[x]^*$. Since $F[x]$ is a PID, it is a UFD. Thus, $f(x)$ is prime in $F[x]$, and hence prime in $R[x]$. This shows that the primeness condition is satisfied. Now, suppose that $f(x), g(x) \in R[x]^*$ are such that f is a proper factor of g. Then either $\deg(f) < \deg(g)$, or there exists some non-unit $u \in R$ such that $g(x) = uf(x)$. Since R is a UFD, it satisfies the ACC, so there can exist no infinite chain of proper factors of u violating the ACC and hence no infinite chain of proper factors of $g(x)$ violating the ACC such that any degree less than or equal to $g(x)$ has infinitely many polynomials of that degree in the chain. Thus, $R[x]^*$ satisfies the ACC, making $R[x]$ a UFD. □

Corollary 3.9.15.1. *If R is a UFD, then so is $R[x_1, \ldots, x_n]$.*

Proof. Follows by induction on n. □

Note. Since multivariable polynomial rings over fields are not PIDs, as we saw earlier in this section, this shows that UFDs need not be PIDs.

3.10 Some Consequences of Factoring

This section, again, follows [Jac09]. We look at two interesting results which follow from our discoveries about polynomial factoring in the previous section. The first is a characterization of when polynomial rings and polynomial rings of functions are the same.

Theorem 3.10.1. *Let F be a field. Then $\mathcal{P}_n(F) \cong F[x_1, \ldots, x_n]$ if and only if F is infinite.*

Proof. If F is finite, then $|\mathcal{P}_n(F)| \leq (n^{|F|})^{|F|}$ and $|F[x_1, \ldots, x_n]| = \infty$, so $F[x_1, \ldots, x_n] \ncong \mathcal{P}_n(F)|$. Now, suppose that F is infinite. There is an obvious homomorphism $\varphi : F[x_1, \ldots, x_n] \to \mathcal{P}_n(F)$ given by $\varphi(f)(a_1, \ldots, a_n) = f(a_1, \ldots, a_n)$. We wish to show that this is an isomorphism. That it is surjective is clear, so we just need to check injectivity. For this, it suffices to show that any non-zero $f(x_1, \ldots, x_n) \in F[x_1, \ldots, x_n]$ has some $a_1, \ldots, a_n \in F$

such that $f(a_1, \ldots, a_n) \neq 0$. We proceed by induction on n. First, suppose that $n = 1$. Then if $f(a) = 0$, a is a root of f. f can have at most $\deg(f) < \infty$ roots, so since $|F| = \infty$ there exists some $a \in F$ such that $f(a) \neq 0$, as required. Now, suppose that the result holds for some $n \geq 1$, and $f \in F[x_1, \ldots, x_{n+1}]$. Then we may write that

$$f(x_1, \ldots, x_{n+1}) = \sum_{k=0}^{r} f_k(x_1, \ldots, x_n) x_{n+1}^k$$

where $f_k \in F[x_1, \ldots, x_n]$, and we may assume without loss of generality that $f_r \neq 0$. By the inductive hypothesis, there exist $a_1, \ldots, a_n \in F$ such that $f_r(a_1, \ldots, a_n) \neq 0$, and hence $f(a_1, \ldots, a_n, x_{n+1}) \in F[x_{n+1}]$ is non-zero. By the case $n = 1$, there is therefore some $a_{n+1} \in F$ such that $f(a_1, \ldots, a_{n+1}) \neq 0$, as required. □

The second concerns the structure of finite subgroups of fields, and becomes quite important in Galois theory.

Theorem 3.10.2. *Any finite subgroup of the multiplicative group F^* is cyclic.*

Proof. Let $G \subset F^*$ be a finite subgroup of the multiplicative group F^*, and let $n = \exp(G)$. Then every element of G must be a root of the polynomial $x^n - 1 \in F[x]$. But $x^n - 1$ can have at most n roots, so there are exactly n elements in G, making it cyclic. □

We can then combine these results to get the following.

Theorem 3.10.3. *Let F be a finite field such that $|F| = q$. Then $\mathcal{P}_n(F) \cong F[x_1, \ldots, x_n]/I$, where $I = (x_1^q - x_1, \ldots, x_n^q - x_n)$.*

Proof. Let $\varphi : F[x_1, \ldots, x_n] \to \mathcal{P}_n(F)$ be the homomorphism from Theorem 3.10.1. It suffices to show that $I = \ker(\varphi)$. Since F is finite, F^* is a cyclic group under multiplication. Thus, $a^q = a$ for any $a \in F$, so certainly $x_k^q - x_k$ are in I. To show that I is generated by the desired polynomials, there are two steps.

First, we show that any $f \in F[x_1, \ldots, x_n]$ of degree strictly less than q in every x_k is not in I. For this, we proceed by induction on n in an identical manner to Theorem 3.10.1. The case $n = 1$ is clear, since a polynomial of degree $< q$ cannot have q roots. The result therefore follows by induction.

Second, we show that any $f \in F[x_1, \ldots, x_n]$ can be written in the form

$$f(x_1, \ldots, x_n) = \sum_{k=1}^{n} f_k(x_1, \ldots, x_n)(x_k^q - x_k) + f_0(x_1, \ldots, x_n)$$

where $f_k \in F[x_1, \ldots, x_n]$ and f_0 is of degree $< q$ in every variable. This implies the desired result, as then $f_0 \in I$ if and only if $f_0 = 0$. It suffices to consider the case of f being a monomial. Consider any monomial of the form $x_1^{j_1} \cdots x_n^{j_n}$. Then for each $1 \leq k \leq n$, there exist $q_k, r_k \in F[x_k]$ such that $x_k^{j_k} = q_k(x_k)(x_k^q - x_k) + r_k(x_k)$, where $\deg(r_k) < q$. Thus,

$$x_1^{j_1} \cdots x_n^{j_n} = (q_1(x_1)(x_1^q - x_1) + r_1(x_1)) \cdots (q_n(x_n)(x_n^q - x_n) + r_n(x_n))$$

Expanding out the expression on the right-hand side, we see that the only term without of factor of $x_k^q - x_k$ for some $1 \leq k \leq n$ is $r_1(x_1) \cdots r_n(x_n)$, which since $\deg(r_k) < q$ is a polynomial of degree $< q$ in every variable, as required. □

3.11 Irreducibility Criteria

In section 3.9 we talked a lot about irreducible polynomials, but I never gave you any tools for recognizing them! This section aims to rectify that, following an identical section in [Lan05][9]. There's not much comment to be made here, it's just three theorems useful for this purpose.

Theorem 3.11.1 (Eisenstein's Criteria)**.** *Let R be a UFD, and $F = FF(R)$. Let $f(x) = a_n x^n + \cdots + a_0 \in R[x]$ be a polynomial of degree at least one. Let $p \in R$ be prime. Then if*

1. *$p \nmid a_n$*
2. *$p \mid a_k$ for all $0 \leq k < n$*
3. *$p^2 \nmid a_0$*

$f(x)$ is irreducible.

Proof. By proposition 3.9.11 and Theorem 3.9.14, we may assume that f is primitive. Suppose $f(x)$ were reducible, say $f(x) = g(x)h(x)$. Let

$$f(x) = a_n x^n + \cdots + a_0 \qquad g(x) = b_m x^m + \cdots + b_0 \qquad h(x) = c_k x^k + \cdots + c_0$$

Then since $p^2 \nmid a_0 = b_0 c_0$, we may assume, without loss of generality, that $p \nmid c_0$ and $p \mid b_0$. Since $p \nmid a_n = b_m c_k$, we conclude that $p \nmid b_m$. We will now show, by induction, that $p \mid b_k$ for all $0 \leq k \leq m$, a contradiction. The case $k = 0$ is done. Suppose it holds for some $k < m$, and every number before that. Then

$$a_{k+1} = b_{k+1}c_0 + b_k c_1 + \cdots + b_0 c_{k+1}$$

where we allow for $c_j = 0$ if necessary. Note that since $f(x)$ is primitive, we may assume that $g(x), c(x)$ are both of degree at least one. Hence, $m < n$, so $p \mid a_{k+1}$. Since $p \nmid c_0$, it follows then by induction that $p \mid b_{k+1}$, as claimed. □

Theorem 3.11.2 (Reduction Criteria)**.** *Let R, R' be integral domains, and $\varphi : R \to R'$ a homomorphism. Let F, F' be the fraction fields of R, R'. Let $f \in R[x]$ be such that $\varphi(f) \neq 0$ and $\deg(\varphi(f)) = \deg(f) \geq 1$. Then if $\varphi(f)$ is irreducible in $F'[x]$, f has no factorization into a product of two degree one or higher polynomials in $R[x]$.*

Proof. Suppose f is reducible in $R[x]$, say $f(x) = g(x)h(x)$, where $\deg(g), \deg(h) \geq 1$. Then $\varphi(f) = \varphi(g)\varphi(h)$, so since $\deg(\varphi(g)) \leq \deg(g)$, and similar with h, we conclude that since $\deg(\varphi(f)) = \deg(f)$, φ preserves the degrees of g, h. Hence, $\varphi(f)$ is reducible in $F'[x]$. □

Theorem 3.11.3 (Integral Root Test)**.** *Suppose R is a UFD and $F = FF(R)$. Let $f(x) = a_n x^n + \cdots + a_0$ be a polynomial in $R[x]$. Let $\alpha \in F$ be a root of f, and write $\alpha = c/d$, where $GCD(c, d) = 1$ (i.e. the GCD are only units). Then $c \mid a_0$ and $d \mid a_n$. Furthermore, if a_n is a unit in R, then $\alpha \in R$.*

[9] Most of the results can also be found in the exercises of [Jac09]

Proof. We get

$$0 = f(\alpha) = a_n(c/d)^n + \cdots + a_0$$

Multiplying both sides through by d^n gives

$$0 = a_n c^n + a_{n-1}c^{n-1}d + \cdots + a_0 d^n$$

Thus, it follows that $c, d \mid a_n c^n + a_0 d^n$. Since $\text{GCD}(c, d) = 1$, the desired result comes from this. □

3.12 Symmetric Polynomials

As we saw in previous sections, multivariable polynomials, while not PIDs, are still UFDs. It's worth asking then whether there's any structure to their factorizations. The answer turns out to be yes, but only for a certain class of multivariable polynomials called *symmetric* polynomials. This section follows similar ones in [Lan05] and [Jac09].

We begin by taking a small detour to talk about *algebraic independence*.

Definition 3.12.1. Let $R \subset S$ be a subring. A set of elements $a_1, \ldots, a_n \in S$ are called algebraically indepenedent over R if there exists no $f \in R[x_1, \ldots, x_n]$ such that $f(a_1, \ldots, a_n)$.

The following result is fairly immediate from this definition, and is left to the reader.

Proposition 3.12.2. *Let $R \subset S$ be a subring, and choose any $a_1, \ldots, a_n \in S$. Then the evaluation homomorphism $\varphi : R[x_1, \ldots, x_n] \to R[a_1, \ldots, a_n]$ is an isomorphism if and only if $a_1, \ldots, a_n$ are algebraically independent.*

This will become relevant later to show a very interesting result. But for now, let's get back to the main topic at hand and work towards defining *symmetric polynomials*.

Definition 3.12.3. Let $f \in R[x_1, \ldots, x_n], \sigma \in S_n$. We define the action of σ on f, denoted $f(\sigma)$, by

$$\sigma(f(x_1, \ldots, x_n)) = f(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$$

f is fixed by σ if $\sigma(f) = f$, and is symmetric if it is fixed by every permutation in S_n.

The set of all symmetric polynomials forms a subring of $R[x_1, \ldots, x_n]$, which we call $\text{Sym}_n(R)$. The main result we build towards is rather surprising, namely that

$$\text{Sym}_n(R) \cong R[x_1, \ldots, x_n]$$

For this, we'll of course need to know what x_k are mapping to. This role will be fulfilled by what we call the *elementary symmetric polynomials*.

Definition 3.12.4. Consider the polynomial $F \in R[x_1, \ldots, x_n][X]$ given by

$$F(X) = (X - x_1)(X - x_2) \cdots (X - x_n)$$

We expand this out, getting an expression of the form

$$F(X) = X^n - s_1 X^{n-1} + \cdots + (-1)^n s_n$$

Then $s_k \in R[x_1, \ldots, x_n]$ are symmetric, and we call s_k the kth elementary symmetric polynomial.

Note. The factors of ± 1 in this definition are arbitrary, and just done to make the expressions in the following proposition a bit nicer. It's also clear from this definition that each x_k is algebraic over $R[s_1, \ldots, s_n]$.

One may ask what these s_k actually look like. This, it turns out, is easy enough to answer.

Proposition 3.12.5. *Let $s_k \in Sym_n(R)$, and let Γ_k be the set of all choices of k numbers from $\{1, \ldots, n\}$. Then*

$$s_k = \sum_{\{a_1, \ldots, a_k\} \in \Gamma_k} x_{a_1} \cdots x_{a_k}$$

Proof. By looking at $F(X)$, we can see that each term in s_k comes from multiplying k different $(-1)x_j$ together. Thus,

$$(-1)^j s_k = \sum_{\{a_1, \ldots, a_k\} \in \Gamma_k} (-1)^j x_{a_1} \cdots x_{a_k}$$

□

We need two more concepts before stating our main result, *homogenaity* and *weight*.

Definition 3.12.6. Let $x_1^{k_1} \cdots x_n^{k_n}$ be a monomial in $R[x_1, \ldots, x_n]$. We define the total degree of this term to be $k_1 + \cdots + k_n$, and the weight to be $k_1 + 2k_2 + \cdots + nk_n$. A polynomial $f \in R[x_1, \ldots, x_n]$ is called homogeneous if all of its terms have the same total degree, and its total degree $t(f)$ and weight $w(f)$ are the maximum total/weight degree of all of its terms.

It's clear from proposition 3.12.5 that the elementary symmetric polynomials are homogeneous, and have total degree k. We can also now, finally, state our main theorems.

Theorem 3.12.7. *Let $f \in R[x_1, \ldots, x_n]$ be symmetric polynomial such that $t(f) = d$. Then there exists $g \in R[x_1, \ldots, x_n]$ such that $w(g) \leq d$ and*

$$f(x_1, \ldots, x_n) = g(s_1, \ldots, s_n)$$

Furthermore, if f is homogeneous, then every monomial in g has weight d.

Proof. We start with induction on n. The result is clear for $n = 1$, since then $s_1 = x$ and $\mathrm{Sym}_n(R) = R[x]$. Now, suppose that the result holds for symmetric polynomials in $n-1$ variables, where $n \geq 2$. We define $s_k^{(0)}$ to be the kth elementary symmetric polynomial in $R[x_1, \ldots, x_n]$ with x_n evaluated to zero. Note that $s_n^{(0)} = 0$, so

$$X(X - x_1) \cdots (X - x_{n-1}) = X(X^{n-1} - s_1^{(0)} X^{n-2} + \cdots + (-1)^{n-1} s_{n-1}^{(0)})$$

Thus, $s_k^{(0)}$ is the kth elementary symmetric polynomial in $R[x_1, \ldots, x_{n-1}]$ for $1 \leq k \leq n-1$. Now, we proceed by induction on d. If $d = 0$, then the result is clear. Suppose the result holds for all symmetric polynomials in $n-1$ variables with total degree $< d$, where $d \geq 1$. Let $f \in \mathrm{Sym}_n(f)$ be such that $t(f) = d$. By the induction on n, there exists some polynomial $g \in R[x_1, \ldots, x_{n-1}]$ of weight $\leq d$ such that

$$f(x_1, \ldots, x_{n-1}, 0) = g(s_1^{(0)}, \ldots, s_{n-1}^{(0)})$$

Note that $t(g) \leq d$ in $R[x_1, \ldots, x_n]$. Thus, we conclude that

$$f_1(x_1, \ldots, x_n) = f(x_1, \ldots, x_n) - g(s_1, \ldots, s_{n-1})$$

is a symmetric polynomial such that $t(f_1) \leq d$. Since $f_1(x_1, \ldots, x_{n-1}, 0) = 0$, we conclude that $x_n \mid f_1$. But f_1 is symmetric, so therefore $x_k \mid f_1$ for all $1 \leq k \leq n$. Therefore, there exists $f_2 \in \mathrm{Sym}_n(R)$ such that

$$f_1(x_1, \ldots, x_n) = x_1 \cdots x_n f_2(x_1, \cdots, x_n)$$

Since $t(f_1) \leq d$, $t(f_2) \leq d - n < d$. Thus, there exists by the inductive hypothesis some $h \in R[x_1, \ldots, x_n]$ such that $w(h) \leq w(f_2)$ and

$$f_2(x_1, \ldots, x_n) = h(s_1, \ldots, s_n)$$

Plugging this back into the above equations we get

$$f(x_1, \ldots, x_n) = x_1 \cdots x_n h(s_1, \ldots, s_n) - g(s_1, \ldots, s_{n-1}) = s_n h(s_1, \ldots, s_n) - g(s_1, \ldots, s_{n-1})$$

Calling $s_n h(s_1, \ldots, s_n) - g(s_1, \ldots, s_{n-1}) = r(s_1, \ldots, s_n)$, we see that $w(r) \leq d$, as required. For the second part of this theorem, we do the same induction. It clearly holds in the base cases, so suppose $f \in \mathrm{Sym}_n(f)$ such that $t(f) = d$ is homogeneous. By the induction on n, there exists some polynomial $g \in R[x_1, \ldots, x_{n-1}]$ with every monomial of weight d such that

$$f(x_1, \ldots, x_{n-1}, 0) = g(s_1^{(0)}, \ldots, s_{n-1}^{(0)})$$

Note that $t(g) = d$ in $R[x_1, \ldots, x_n]$. Thus, we conclude that

$$f_1(x_1, \ldots, x_n) = f(x_1, \ldots, x_n) - g(s_1, \ldots, s_{n-1})$$

is either zero (in which case we're done) or a symmetric polynomial such that $t(f_1) = d$. Since $f_1(x_1, \ldots, x_{n-1}, 0) = 0$, we conclude that $x_n \mid f_1$. But f_1 is symmetric, so therefore $x_k \mid f_1$ for all $1 \leq k \leq n$. Therefore, there exists a homogeneous $f_2 \in \mathrm{Sym}_n(R)$ such that

$$f_1(x_1, \ldots, x_n) = x_1 \cdots x_n f_2(x_1, \cdots, x_n)$$

Since $t(f_1) = d$, $t(f_2) = d - n$. Thus, there exists by the inductive hypothesis some $h \in R[x_1, \ldots, x_n]$ such that every monomial in h has weight $d - n$ and

$$f_2(x_1, \ldots, x_n) = h(s_1, \ldots, s_n)$$

Plugging this back into the above equations we get

$$f(x_1, \ldots, x_n) = x_1 \cdots x_n h(s_1, \ldots, s_n) - g(s_1, \ldots, s_{n-1}) = s_n h(s_1, \ldots, s_n) - g(s_1, \ldots, s_{n-1})$$

Calling $s_n h(s_1, \ldots, s_n) - g(s_1, \ldots, s_{n-1}) = r(s_1, \ldots, s_n)$, we see that the weight of every monomial in r is d, as required. □

Theorem 3.12.8. *The elementary symmetric polynomials are algebraically independent over R.*

Proof. The result is clear for the case $n = 1$, so we proceed by induction. Let $n \geq 2$, suppose the result holds for $< n$, and that there existed some $f \in R[x_1, \ldots, x_n]$ such that $f(s_1, \ldots, s_n) = 0$. In particular, choose f to have a minimal (non-zero) total degree, call it $t(f) = d$. Write

$$f(x_1, \ldots, x_n) = f_0(x_1, \ldots, x_{n-1}) + \cdots + f_m(x_1, \ldots, x_{n-1})x_n^m$$

where $f_k \in R[x_1, \ldots, x_n]$ and $f_m \neq 0$. We can note that, using the notation of the proof of Theorem 3.12.7

$$0 = f(s_1^{(0)}, \ldots, s_{n-1}^{(0)}, 0) = f_0(s_1^{(0)}, \ldots, s_{n-1}^{(0)})$$

By the inductive hypothesis, $s_1^{(0)}, \ldots, s_{n-1}^{(0)}$ are algebraically independent over $R[x_1, \ldots, x_{n-1}]$, so $f_0 = 0$. But then if $f \neq 0$

$$f(x_1, \ldots, x_n) = x_n(f_1(x_1, \ldots, x_{n-1}) + \cdots + f_m(x_1, \ldots, x_{n-1})x_n^{m-1})$$

so $f_1(x_1, \ldots, x_{n-1}) + \cdots + f_m(x_1, \ldots, x_{n-1})x_n^{m-1}$ is a polynomial of lower total degree which evaluates to zero on $s_1, \ldots, s_n$, a contradiction. Hence, $f = 0$ as required. □

This of course gives the following immediate result.

Corollary 3.12.8.1. $R[s_1, \ldots, s_n] \cong R[x_1, \ldots, x_n]$.

3.13 Complex Numbers and Quaternions*

The topics in this section will be pretty much completely irrelevant for the rest of the text, the author just gave a lecture on them once that they liked and doesn't want that work to go to waste. The material contained here is pulled from a combination of [Jac09], [Gub21], and the author's own head. It assumes basic knowledge of the complex numbers.

The real numbers, from an analytic perspective, are wonderful. They are however, from an algebraic perspective, terrible. Why? Well, in algebra we're usually in the business of solving for the roots of polynomials. And it turns out that a lot of very simple polynomials in $\mathbb{R}[x]$ have no roots in $\mathbb{R}[x]$, and in fact are irreducible. To fix this, we're going try defining a new ring which adds solutions to polynomials to $\mathbb{R}$. Let's start by adding a solution to the simplest polynomial in $\mathbb{R}[x]$ with no roots, $x^2 + 1$. To do this, we define our new ring to be

$$C = \frac{\mathbb{R}[x]}{(x^2 + 1)}$$

What we've done here is sort of a sleight of hand. If we let $I = (x^2 + 1)$, then we can see that $x + I \in C$. Thus, we can evaluate $f(x) = x^2 + 1$ in C at $x + I$, which gives the following

$$f(x + I) = (x + I)^2 + (1 + I) = (x^2 + 1) + I = 0$$

The root we've added to f is exactly $x + I$!. It's not too hard to see that $-x + I$ is also a (distinct) root of f in this new ring. In fact, it turns out that *all* the roots of *every* polynomial in $\mathbb{R}[x]$ are contained in C. The reason for this is the following very simple theorem.

Theorem 3.13.1. $C \cong \mathbb{C}$.

Proof. Note that since $x^2 = 1$ in C, any element in C has a unique representation of the form $a + bx$, where $a, b \in \mathbb{R}$. It is then a quick verification that $\varphi(a + bx) = a + bi$ is a well-defined isomorphism. □

One could choose to define $\mathbb{C}$ in this manner, as the field $\mathbb{R}[x]/(x^2 + 1)$. In an algebraic context, this is actually a quite intuitive way of defining $\mathbb{C}$. The statement above about finding roots for any polynomial also falls out of this result, as we have (from many different field) the following.

Theorem 3.13.2 (The Fundamental Theorem of Algebra)**.** *Every $f \in \mathbb{R}[x]$ has, including multiplicity*[10]*, n roots in $\mathbb{C}$.*

Great, we've added all the solutions to polynomials we could ever want. But we don't stop here, because there's a new problem. As you would learn almost immediately in any class covering the complex numbers, $\mathbb{C}$ is just $\mathbb{R}^2$ endowed with a multiplication. So can we in turn endow $\mathbb{C}^2$, or equivalently $\mathbb{R}^4$, with a multiplication? The answer is yes, if we're willing to give up on that multiplication being commutative.

Definition 3.13.3. The *quaternions*, denoted $\mathbb{H}$, are $\mathbb{C}^2$ with the standard vector addition and a multiplication given by

$$(a, b) \cdot (c, d) = (ac - b^* d, da + bc^*)$$

There are many useful properties and representations of the quaternions. Proving them is mostly just very tedious symbol pushing, so we simply list them below.

Proposition 3.13.4. *$\mathbb{H}$ is a non-commutative division ring.*

Proposition 3.13.5. *The following spaces are all isomorphic.*

1. $\mathbb{H}$
2. $\mathbb{R}[i, j, k]/I$, *where* $I = (i^2 + 1, j^2 + 1, k^2 + 1, ij + ji, k + ji, jk + kj, i + kj, ki + ik, j + ik)$
3. *The subset of $\mathbb{C}^{2\times 2}$ consisting of matrices of the form*

$$\begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix}$$

where $a, b \in \mathbb{C}$, and $*$ is complex conjugation.

Furthermore, in the standard form of these isomorphisms, the following elements are equivalent (where $a, b, c, d \in \mathbb{R}$)

1. $(a + bi, c + di)$

[10] This is defined in the usual manner from grade school.

2. $a + bi + cj + dk$

3.
$$\begin{pmatrix} a+bi & c+di \\ -c+di & a-bi \end{pmatrix}$$

The latter of these forms also gives us the inverse of any non-zero quaternion, using the formula for the inverse of a $2x2$ matrix.

Proposition 3.13.6. *If $(a, b) \in \mathbb{H}$ is non-zero, then $(a, b)^{-1} = (|a|^2 + |b|^2)^{-1}(a^*, -b)$.*

We call this factor $|a|^2 + |b|^2$ the *norm* of the quaternion, and denote it $N((a, b))$. One can note that this is the determinant of the matrix

$$\begin{pmatrix} a+bi & c+di \\ -c+di & a-bi \end{pmatrix}$$

Thus, N is a multiplicative homomorphism and $N(a+bi+cj+dk) = a^2+b^2+c^2+d^2$. We call $(a^*, -b)$ the *conjugation* of the quaternion, and denote it $(a, b)^*$. Thus, the above proposition could be compressed to saying that for $a \in \mathbb{H}$, $a^{-1} = \frac{a^*}{N(a)}$, just like the inverses in $\mathbb{C}$. One can also check that $aa^* = N(a)$, and that the complex conjugate is a field automorphism of $\mathbb{H}$.

All these results are well and good, but to justify doing all of this let's look at some useful applications of quaternions. All of classical physics can technically be formulated in terms of quaternions, it's just a really bad way to do it. Except for exactly one area, **rotations**.

Proposition 3.13.7. *Every quaternion $h \in \mathbb{H}$ can be written uniquely in the form*

$$h = a(\cos(\theta) + \mathbf{n}\sin(\theta))$$

where $\theta, a \in \mathbb{R}$, $\mathbf{n} = n_1 i + n_2 j + n_3 k$, and $n_1^2 + n_2^2 + n_3^2 = 1$.

Note that $\mathbf{n}$ is essentially a unit vector in $\mathbb{R}^3$, and that a acts like a magnitude. Thus, numbers in $\mathbb{H}$ with real component 0, which we denote $\mathrm{Im}(\mathbb{H})$, encode $\mathbb{R}^3$. This connects with rotations in the following manner.

Proposition 3.13.8. *Let $q = \cos(\theta) + \mathbf{n}\sin(\theta) \in \mathbb{H}$, and let $\mathbf{h} \in Im(\mathbb{H})$. Then the map $\mathbf{h} \mapsto q\mathbf{h}q^*$ is the rotation of $\mathbf{h}$ by an angle 2θ about the axis collinear to $\mathbf{n}$ in $\mathbb{R}^3$.*

This gives us an efficient way to store and compute rotations! We can also, in a similar manner, use quaternions to compute the cross and dot product. In fact, looking at vectors in $\mathbb{R}^3$ as quaternions

$$\mathbf{u} \cdot \mathbf{v} = -\frac{\mathbf{uv} + \mathbf{vu}}{2} \qquad \mathbf{u} \times \mathbf{v} = \frac{\mathbf{uv} - \mathbf{vu}}{2}$$

Quaternions also have numerous uses in number theory, many of which are outlined in [Gub21].

There is, of course, one remaining question. Can we pull this trick again, and endow $\mathbb{H}^2$ with a multiplication? The answer is yes, but the result won't be a ring and will instead be something called a $\mathbb{R}$ (or $\mathbb{C}$)-algebra. Essentially, you can multiply numbers in $\mathbb{H}^2$, but that multiplication won't be associative. This trend continues, with each step up the ladder losing more and more nice properties that multiplication could have. If you'd like to learn more about this, see [Gub21].

3.14 Chinese Remainder Theorem*

The Chinese Remainder Theorem is one of the most fundamental theorems in ring theory, and yet fit nowhere anywhere else in this chapter. Nor is it used again for the remainder of this book, except perhaps in the final chapter. I put it here for lack of a better place, but despite its "optional" marking I would highly recommend going over this section. The content of this section is based on lectures given by Dr. Kalle Karu at UBC.

Let's start with some preliminaries.

Definition 3.14.1. Let R be a ring and $\{I_j\}_{j=1}^n$ a set of ideals of R. We define

1. $$\sum_{j=1}^{n} I_j = \Big\{ \sum_{j=1}^{n} f_j \mid f_j \in I_j \Big\}$$
2. $$\prod_{j=1}^{n} I_j = \Big\{ \sum_{k=1}^{m} \prod_{j=1}^{n} f_{j,k} \mid f_{j,k} \in I_j, m \in \mathbb{Z}^+ \Big\}$$

These are called the sum and product ideals respectively.

It is not too hard to show that, like the names suggest, the sum and product ideals are ideals in R (the latter is only guaranteed to be an ideal when R is commutative). It is also clear that $\prod_{j=1}^n I_j \subset \bigcap_{j=1}^n I_j$. The Chinese Remainder theorem first tells us sufficient conditions for these two expressions to be equal, namely the condition of ideals being coprime.

Definition 3.14.2. Two ideals $I, J \subset R$ are coprime if $I + J = R$.

Theorem 3.14.3 (Chinese Remainder Theorem (CRT) I)**.** *Let $I_1, \ldots, I_n \subset R$ be a collection of pairwise coprime ideals in a commutative ring. Then*

$$\prod_{j=1}^{n} I_j = \bigcap_{j=1}^{n} I_j$$

Proof. It suffices to show that $\prod_{j=1}^n I_j \supset \bigcap_{j=1}^n I_j$. First, suppose that $n = 2$. Pick any $x \in I_1 \cap I_2$. Since $I_1 + I_2 = R$, there exist some $y_1, y_2 \in I_1, I_2$ such that $y_1 + y_2 = 1$. Therefore,

$$x = x(y_1 + y_2) = y_1 x + x y_2 \in I_1 I_2$$

as required. Now, suppose $n \geq 2$. Set $\prod_{j=1}^{n-1} I_j = I$. By the case $n = 2$, it suffices to show that I, I_n are coprime. Since I_k, I_n are pairwise coprime for each $1 \leq k \leq n-1$, we can find $x_k \in I_k, y_k \in I_n$ such that $x_k + y_k = 1$. Then $\prod_{k=1}^{n-1} x_k \in I$,

$$\sum_{k=1}^{n-1} y_k \prod_{j=k+1}^{n-1} x_j \in I_n$$

where we say that $\prod_{j=k+1}^{n-1} x_j = 1$ when $k = n-1$, and one can verify

$$\prod_{k=1}^{n-1} x_k + \sum_{k=1}^{n-1} y_k \prod_{j=k+1}^{n-1} x_j = 1$$

Thus, $I + I_n = (1) = R$ as claimed. □

Corollary 3.14.3.1 (Chinese Remainder Theorem II)**.** *Endow the cartesian product of rings with a ring structure via elementwise operations. Let $I_1, \ldots, I_n \subset R$ be ideals in an arbitrary ring. Let $q : R \to R/I_1 \times \cdots \times R/I_n$ be the ring homomorphism induced by the quotient maps, that is the one given by*

$$q(x) = (x \bmod I_1, \ldots, x \bmod I_n)$$

Then q is surjective if and only if the I_k are pairwise coprime.

Proof. Suppose q is surjective. Pick any $1 \leq k \leq n$, and $j \neq k$. Then there exists some $x \in R$ such that

$$q(x) = (\ldots, 1, 0, 0, \ldots)$$

where the 1 is in the kth position. Then $x \in I_j$ and $\exists y \in I_k$ such that $x = y + 1$, so $x - y = 1 \Rightarrow I_j + I_k = (1) = R$. Thus, all the ideals are pairwise coprime.

Now, suppose that all the ideals are pairwise coprime. Set $I = \prod_{j=1}^{n-1} I_j$. By the proof of CRT I, we can find $x \in I, y \in I_n$ such that $x + y = 1$. Then $q(x) = (0, 0, \ldots, 1)$, and it follows by a symmetric argument to this that q is surjective. □

Corollary 3.14.3.2 (Chinese Remainder Theorem III)**.** *Let $I_1, \ldots, I_n \subset R$ be coprime ideals in an arbitrary ring. Then*

$$\frac{R}{\bigcap_{i=1}^{n} I_i} = R/I_1 \times \cdots \times R/I_n$$

Also, if R is commutative then

$$\frac{R}{\prod_{i=1}^{n} I_i} = R/I_1 \times \cdots \times R/I_n$$

Proof. Let Let $q : R \to R/I_1 \times \cdots \times R/I_n$ be the ring homomorphism induced by the quotient maps, as defined in CRT II. By CRT I/II, it suffices to show that $\ker(q) = \bigcap_{i=1}^{n} I_i$. But this is immediate. □

For an interesting example of CRT in action, I suggest you look into the history of the theorem, particularly its original form in terms of $\mathbb{Z}$.

Part II

Linear Algebra

Chapter 4

Modules

4.1 Basics Definitions

Modules, in a broad sense, are simply generalizations of vector spaces to be over arbitrary rings rather than fields. We begin their study here, following (loosely) similar explanations from [Jac09] and [Lan05].

Definition 4.1.1. Let R be a ring. A left R-module M is an Abelian group $(M, +, 0)$ together with a scalar multiplication operation $\cdot : R \times M \to M$ satisfying the following axioms for all $x, y \in R, \underline{v}, \underline{u} \in M$

1. $x \cdot (\underline{v} + \underline{u}) = x \cdot \underline{v} + x \cdot \underline{u}$
2. $(x + y) \cdot \underline{v} = x \cdot \underline{v} + y \cdot \underline{v}$
3. $x \cdot (y \cdot \underline{v}) = (xy) \cdot \underline{v}$
4. $1 \cdot \underline{v} = \underline{v}$

We also have right R-modules, which are defined similarly.

Definition 4.1.2. Let R be a ring. A right R-module M is an Abelian group $(M, +, 0)$ together with a scalar multiplication operation $\cdot : M \times R \to M$ satisfying the following axioms for all $x, y \in R, \underline{v}, \underline{u} \in M$

1. $(\underline{v} + \underline{u}) \cdot x = \underline{v} \cdot x + \underline{u} \cdot x$
2. $\underline{v} \cdot (x + y) = \underline{v} \cdot x + \underline{v} \cdot y$
3. $(\underline{v} \cdot x) \cdot y = \underline{v} \cdot (xy)$
4. $\underline{v} \cdot 1 = \underline{v}$

A quick note about notation before we move on : like with everything else in algebra, we generally drop the $\cdot$ from our scalar multiplication expressions and just write $x\underline{v}$. We will

also, in this text, use the convention of underlining symbols which represent module elements. This is not standard, and most other texts will have no particular convention in this regard.

At first glance, left and right R-modules seem like the exact same thing, and if R is commutative they in fact are the same thing. But when R is not commutative we get complications. To understand why, we need to take a diversion into *endomorphisms of Abelian groups* and *anti-morphisms*. Let's start by examining the structure of a left R-module M.

Proposition 4.1.3. *Suppose M is a left R-module. For each $x \in R$, let $\varphi_x : M \to M$ be the map given by $\varphi_x : \underline{v} \to x\underline{v}$. Then the map $f : x \mapsto \varphi_x$ is a ring homomorphism from R into* $\mathrm{End}(M)$*, the ring of endomorphisms of M as an Abelian group.*

Proof. That φ_x is an endomorphism of M is given by the first axiom in definition 4.1.2. To check that $f : x \mapsto \varphi_x$ is a ring homomorphism, we first need to give a ring structure to $\mathrm{End}(M)$. Pick any $\varphi, \psi \in \mathrm{End}(M)$ and $\underline{v} \in M$. It is not too hard to see that the operations

$$(\varphi + \psi)(\underline{v}) = \varphi(\underline{v}) + \psi(\underline{v}) (\varphi \cdot \psi)(\underline{v}) = (\varphi \circ \psi)(\underline{v})$$

put a ring structure on $\mathrm{End}(M)$. That f is a homomorphism is then guaranteed by axioms 2-4 of definition 4.1.2. □

Of course, the above result also tells us that any homomorphism from R to $\mathrm{End}(M)$ will give us a left R-module structure. Thus, we could have defined a left R-module M as a commutative group M with ring homomorphism $R \to \mathrm{End}(M)$. This is where right R-modules differ. The map $\underline{v} \mapsto \underline{v} \cdot x$ will still be an endomorphism of M. However, with the ring structure we've given to $\mathrm{End}(M)$, we would get that $f(xy) = f(y)f(x)$, not $f(x)f(y)$. This type of function is called an *anti-morphism*, and is actually quite similar to homomorphisms. In fact, an analogous result to proposition 4.1.3 holds for right R-modules using anti-morphisms. However, anti-morphisms are not homomorphisms (in general), so this shows that left and right R-modules are not necessarily the same. One may note that they are the same if R is commutative, hence the earlier statement that left and right R-modules are no different for commutative R.
For the rest of this chapter, we work with left R-modules. Keep in mind that analogous results will hold for right R-modules in almost all cases, if you wish to see these results explicitly see [Jac09]. With this in mind, let us give some basic definitions of module theory.

Definition 4.1.4. Let M be a left R-module. A submodule $N \subset M$ is an additive subgroup of M such that, $RN \subset N$, where

$$RN = \{r\underline{v} \mid r \in R, \underline{v} \in N\}$$

Like with ideals, we also have the following two constructions which are also subgroups.

Proposition 4.1.5. *Let $\{M_i\}_{i \in I}$ be a collection of submodules of a left R-module M. Then*

1. $\bigcap_{i \in I} M_i$

2.

$$\sum_{i\in I} M_i = \Big\{ \sum_{i,finite} \underline{v}_i \mid \underline{v}_i \in M_i \Big\}$$

are both sub-modules of M.

The former of these allows us to make the following definition.

Definition 4.1.6. Let $S \subset M$, where M is a left R-module. Then the submodule generated by S, denoted $\mathrm{Span}_R(S)$, is the intersection of all submodules of M containing S.

Again, like for subgroups and ideals, we have a simple way of explicitly writing out these generated sub-modules.

Proposition 4.1.7. *Let $S \subset M$, where M is a left R-module. Then*

$$\mathrm{Span}_R(S) = \Big\{ \sum_{i=1}^{n} r_i \underline{v}_i \mid r_i \in R, \underline{v}_i \in S, n \in \mathbb{Z}^+ \Big\}$$

Before moving on, let's list a pair of facts about modules that would be no fun to prove.

Proposition 4.1.8. *Suppose that M is a left R-module, and $N \subset M$ a submodule. Then*

1. *$0\underline{v} = \underline{0}$ and $(-1)\underline{v} = -\underline{v}$, for any $\underline{v} \in M$.*
2. *N is an additive subgroup of M.*

Next, we move to talking about module homomorphisms.

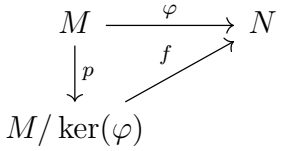
Definition 4.1.9. Let M, N be left R-modules. A module homomorphism $\varphi : M \to N$ is a homomorphism of additive groups satisfying, for any $x \in R, \underline{v} \in M$, $\varphi(r\underline{v}) = r\varphi(\underline{v})$. We denote the set of all module homomorphisms from M to N by $\mathrm{Hom}_R(M, N)$.

Note. $\mathrm{Hom}_R(M, N)$ is itself a left R-module. You will also here module homomorphisms be called R-linear maps.

The kernel of a module homomorphism is just the kernel of the underlying homomorphism of additive groups. It is easy to show that this kernel (and the image of a module homomorphism) is a submodule as well. One can also check that, if N is a submodule of M, then M/N is itself a left R-module, with operations inherited in the normal way from N.

At this point, I'm going to do something quite interesting. In the previous sections, I proved the fundamental theorems of homomorphisms. Here, I'm just going to state them.

Theorem 4.1.10 (First Fundamental Theorem of Module Homomorphisms)**.** *Let $\varphi \in \mathrm{Hom}_R(M, N)$ be a module homomorphism. Then the natural projection map $p : M \to M/\ker(\varphi)$ is a module homomorphism, and the map $f : M/\ker(\varphi) \to \mathrm{Im}(\varphi)$ given by $f : \underline{v} + \ker(\varphi) \to \varphi(\underline{v})$ is a well-defined module isomorphism. Finally, the following diagram commutes.*



Theorem 4.1.11 (Second Fundamental Theorem of Module Homomorphisms)**.** *Let* $\varphi \in \mathrm{Hom}_R(M, N)$ *be a surjective module homomorphism. Then*

1. *An additive subgroup* $S \subset M$ *containing* $\ker(\varphi)$ *is a submodule if and only if* $\varphi(S)$ *is a submodule.*
2. *The map* $S \mapsto \varphi(S)$ *of submodules of* M *containing* $\ker(\varphi)$ *is a bijection onto submodules of* N*.*
3. *If* $M' \subset M$ *is a submodule containing* $\ker(\varphi)$*, then* $M/M' \cong N/\varphi(M')$*.*

Corollary 4.1.11.1. *Suppose* $K \subset N$ *are both submodules of a left* R*-module* M*. Then*

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

Theorem 4.1.12 (Third Fundamental Theorem of Module Homomorphisms)**.** *Let* N, K *be submodules of a left* R*-module* N*. Then*

$$\frac{N}{N \cap K} \cong \frac{N + K}{K}$$

I don't bother with the proofs here for a good reason : these theorems are again essentially identical to those found in section 2.4 and section 3.4. At this point, you should be able to do them on your own, or at the very least believe them when you see them. If you do not feel that you've reached this point yet, then I would suggest reading those two sections again.

4.2 Free Modules and Bases

We take on now the vitally important task of generalizing bases from vector spaces to modules. To do so, we take a synthesis of similar sections in [Jac09] and [Lan05], along with some insights from course notes [Bad10] and linear algebra on vector spaces [Rot07].

Definition 4.2.1. Let M be a left R-module, and $S \subset M$. We call S

1. Linearly independent if
$$\sum_{\underline{v} \in S} a_{\underline{v}} \underline{v} = 0 \Rightarrow a_{\underline{v}} = 0, \forall \underline{v} \in S$$
2. Linearly dependent if it is not linearly independent.
3. A basis if it is linearly independent and $\mathrm{Span}_R(S) = M$.

These definitions are identical to those we use in vector spaces.

Note. In expressions such as

$$\sum_{\underline{v}\in S} a_{\underline{v}}\underline{v} = 0 \Rightarrow a_{\underline{v}} = 0, \forall \underline{v} \in S$$

we always assume that only finitely many terms have a non-zero $a_{\underline{v}}$, and ignore those with zero. Indeed, the expression is not well-defined otherwise. This just gives us a compact way to represent all finite linear combinations of elements in S.

Many of the nice properties you're used to from vector space bases carry over to modules as well.

Proposition 4.2.2. *Let M be a free left R-module with basis $V = \{\underline{v}_i\}_{i\in I}$. Then each $\underline{u} \in M$ can be written in a unique way in the form*

$$\underline{u} = \sum_{i\in I} a_i\underline{v}_i$$

where $a_i \in R$.

Proof. The existence is simply the statement that V is spanning. For uniqueness, suppose that

$$\sum_{i\in I} a_i\underline{v}_i = \sum_{i\in I} v_i\underline{v}_i$$

Rearranging, we get

$$\sum_{i\in I}(a_i - b_i)\underline{v}_i = \underline{0}$$

which by linear independent implies that $a_i = b_i$ for every $i \in I$. □

Proposition 4.2.3. *Let M be a free left R-module with basis $V = \{\underline{v}_i\}_{i\in I}$. Let N be any other left R-module, and for each $i \in I$ choose some $\underline{u}_i \in N$. Then there exists a unique $\varphi : \mathrm{Hom}_R(M, N)$ such that $\varphi(\underline{v}_i) = \underline{u}_i$, for all $i \in I$.*

Proof. By the previous proposition, and $\underline{v} \in V$ can be written uniquely in the form

$$\underline{v} = \sum_{i\in I} a_i\underline{v}_i$$

Thus, by linearity the only way φ could be defined is

$$\varphi(\underline{v}) = \sum_{i\in I} a_i\underline{u}_i$$

This gives us that φ is unique and well-defined. Checking that it is a module homomorphism is simple, and left to the reader. □

Corollary 4.2.3.1. *If M, N are left R-modules with bases V, U such that $|V| = |U|$, then $M \cong N$.*

In light of that last result, it is natural to ask whether modules have a unique basis cardinality (or for finite cases, basis size). The answer, it turns out, is sufficiently strange that we'll spend the rest of this section answering it.

Let's start by simplifying our terminology. Let I be an arbitrary set, and R a ring. We denote $\prod_{i\in I} R$ by R^I. We can endow a left R-module structure on this in the following way.

$$(a_i)_{i\in I} + (b_i)_{i\in I} = (a_i + b_i)_{i\in I} \qquad\qquad c \cdot (a_i)_{i\in I} = (ca_i)_{i\in I}$$

We denote the sequence with a one in the ith position and zeroes everywhere else by $\underline{e}_i \in R^I$. It is not hard to check that $\mathcal{B} = \{\underline{e}_i\}_{i\in I}$ is a basis for R^I, we call this the *standard basis*. By corollary 4.2.3.1, any free left R-module is isomorphic to R^I for some set I. Thus, we can reduce our study of free left R-modules to just the study of those R^I.

In light of this, we can carry over more results we know from basic linear algebra. In particular, in "finite-dimensional" spaces we can represent linear maps as matrices (relative to a pair of basis chosen), with the process being identical to that done for vector spaces. As such, the following result probably shouldn't be too surprising.

Theorem 4.2.4. *Let $n, m \in \mathbb{N}$. Then $R^n \cong R^m$ if and only if there exists a pair of matrices $A \in M_{n,m}(R), B \in M_{m,n}(R)$ such that $AB = \mathrm{Id}_n, BA = \mathrm{Id}_m$.*

Proof. First, suppose that there exists an isomorphism $\varphi : R^n \to R^m$. Let $\{\underline{e}_i\}_{i=1}^n, \{\underline{x}_j\}_{j=1}^m$ be the standard bases for R^n, R^m respectively. Then since φ is an isomorphism, $\varphi(\{\underline{e}_i\}_{i=1}^n)$ is a basis for R^m (this is not hard to check). We'll write $\underline{y}_i = \varphi(\underline{e}_i)$. Then for each $1 \le i \le n$, there exist unique $a_{i,j} \in R$ such that

$$\underline{y}_i = \sum_{j=1}^{m} a_{i,j}\underline{x}_j$$

Similarly, for each $1 \le j \le m$ we can find unique $b_{j,i}$ such that

$$\underline{x}_j = \sum_{i=1}^{n} b_{j,i}\underline{y}_i$$

We write $A = (a_{i,j}), B = (b_{j,i})$. We'll show that these are the desired matrices. First, we get

$$(AB)_{i,j} = \sum_{k=1}^{m} a_{i,k}b_{k,j}$$

Now, since

$$\underline{y}_i = \sum_{j=1}^{m} a_{i,j}x_j = \sum_{j=1}^{m}\sum_{k=1}^{n} a_{i,j}b_{j,k}\underline{y}_i = \sum_{k=1}^{n}\Big(\sum_{j=1}^{m} a_{i,j}b_{j,k}\Big)\underline{y}_k$$

By the uniqueness of the $a_{i,j}$ in the original expression, it follows that

$$(AB)_{i,j} = \sum_{k=1}^{m} a_{i,k}b_{k,j} = \delta_{ij} \Rightarrow AB = \mathrm{Id}_n$$

We also get $BA = \mathrm{Id}_m$ by the same argument.

Now, suppose that we have such matrices A, B. Then B is an R-linear map from R^n to R^m, which the existence of A shows is invertible and hence an isomorphism. □

In the case of commutative rings R, this result gives what we'd expect due to the following lemma.

Lemma 4.2.5. *Suppose R is a commutative ring and $A, B \in M_n(R)$. Then $AB = \mathrm{Id}_n \Rightarrow BA = \mathrm{Id}_n$.*

Proof. This follows from the properties of the determinant. Indeed, we get

$$\det(BA) = \det(B)\det(A) = \det(A)\det(B) = \det(AB) = 1$$

so BA is invertible. Then we note

$$BA = B(\mathrm{Id}_n)A = B(AB)A = (BA)^2$$

Applying $(BA)^{-1}$ on both sides, which we now know exists, gives the desired result. □

Note. This result **does not hold** if R is not commutative.

Corollary 4.2.5.1. *Suppose $n, m \in \mathbb{N}$ and R is commutative. Then $R^n \cong R^m$ if and only if $n = m$.*

Proof. The direction assuming $n = m$ is obvious, so instead assume that $R^n \cong R^m$. Assume, without loss of generality, that $n < m$. Using the same notation as in the proof of Theorem 4.2.4, we instead define $A, B \in M_m(R)$ by

$$A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,m} \\ a_{2,1} & \cdots & a_{2,m} \\ \vdots & \vdots & \vdots \\ a_{n,1} & \cdots & a_{m,m} \\ 0 & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \qquad B = \begin{pmatrix} b_{1,1} & \cdots & a_{1,n} & 0 & \cdots & 0 \\ b_{2,1} & \cdots & b_{2,n} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{m,1} & \cdots & b_{m,n} & 0 & \cdots & 0 \end{pmatrix}$$

Then by the calculation in Theorem 4.2.4, we'd still get $BA = \mathrm{Id}_n$, but $AB \neq \mathrm{Id}_n$, violating lemma 4.2.5. □

The above note is quite important here, as there are non-commutative rings R where $R^n \cong R^m$ and $n \neq m$. I'll leave it to the interested reader to look up counterexamples, as they don't tend to be one-line things. In general, we call any ring satisfying the above corollary an *invariant basis number* (IBN) ring.

Perhaps the most remarkable fact out of all of this is that modules with infinite bases are more well-behaved than finite ones! Indeed, if we replace n, m with infinite sets, then the above corollary becomes true for any ring. More precisely, we have the following.

Theorem 4.2.6. *Suppose I, J are sets such that I is infinite. Then $R^I \cong R^J$ if and only if $|I| = |J|$.*

Proof. Again, it suffices to show that $R^I \cong R^J \Rightarrow |I| = |J|$. First, we'll show that J is infinite. Indeed, suppose that J were finite. Let $\varphi : R^I \to R^J$ be a module isomorphism, $\underline{x}_i$ be the images of the standard basis elements of R^I under φ, and $\underline{y}_j$ be the standard basis elements of R^J. Since φ is an isomorphism, $\{\underline{x}_i\}_{i\in I}$ is a basis of R^J. Thus, each $\underline{y}_j$ can be written as a linear combination of finitely many $\underline{x}_i$. But since J is finite, this would imply that finitely many $\underline{x}_i$ span R^J, and hence that $\{\underline{x}_i\}_{i\in I}$ is not linearly independent. This is impossible, and hence J must be infinite.

Now, we can assume that J is also infinite. For each $j \in J$, let $U_j \subset I$ be a finite subset such that $\underline{y}_j$ can be expressed as a linear combination of vectors in $\{\underline{x}_i\}_{i\in U_j}$. Then since $\{\underline{y}_j\}_{j\in J}$ is a basis, we must get

$$\bigcup_{j\in J} U_j = I$$

Therefore, $|J| \leq |\bigcup_{j\in J} U_j| \leq |I|$. A symmetric argument shows that $|I| \leq |J|$, completing the proof. □

Note. If you're not comfortable playing around with set cardinalities like this, I suggest looking at the preliminaries in [Rot07].

If a module has a basis (this is not guaranteed), then we call the cardinality of this basis its *rank*. Note that finite ranks, when they exist, may not be unique for non-IBN rings.

4.3 Direct Sums and Products

This section primarily follows the notes of [Sil23], with some ideas from [Lan05] and [Rot07]. We start by defining *direct sums*.

Definition 4.3.1. Let $\{M_i\}_{i\in I}$ be a collection of left R-modules. The direct sum of these modules, denoted $\bigoplus_{i\in I} M_i$, is the set of sequences in $\prod_{i\in I} M_i$ with finite support (i.e. finitely many non-zero values) with addition defined element-wise and multiplication by

$$r(\underline{m}_i)_{i\in I} = (r\underline{m}_i)_{i\in I}$$

These have a couple properties which may be verified without too much difficulty.

Proposition 4.3.2. *Let $\{M_i\}_{i\in I}$ be a collection of left R-modules. Then*

1. *Each M_i can be embedded in $M = \bigoplus_{i\in I} M_i$ via the canonical set embedding. That is, calling our inclusions $\iota_i : M_i \to M$, we define that $\iota_i(\underline{m})$ to be the sequence with zeroes everywhere except having $\underline{m}$ in the ith position.*
2. *Suppose $\{B_i\}_{i\in I}$ is a collection of bases for each M_i. Then $\bigcup_{i\in I} \iota_i(B_i)$ is a basis for M.*
3. *If $n, m \in \mathbb{N}$, then $R^n \oplus R^m \cong R^{n+m}$.*

4. $(M_1 \oplus M_2) \oplus M_3 \cong M_1 \oplus (M_2 \oplus M_3)$

There is a (sometimes) closely related notion that we've encountered, called the *internal sum.*

Definition 4.3.3. Let $\{N_i\}_{i\in I} \subset M$ be sub-modules of a left R-module M. Then we define the internal sum of these to be

$$\sum_{i\in I} N_i = \mathrm{Span}_R\left(\bigcup_{i\in I} N_i\right) = \{\textit{finite sums of elements from the } N_i\}$$

Let $\varphi : \bigoplus_{i\in I} N_i \to \sum_{i\in I} N_i$ be the homomorphism defined by being the inclusion on each N_i. This sum is called *direct* if φ is an isomorphism.

Ideally, we'd like all our sums to be direct, as breaking a module down into a direct sum of simpler modules makes it much easier to work with. This, of course, isn't the case, but we do have simple rules for detecting when internal sums are and are not direct.

Theorem 4.3.4. *Let $\{N_i\}_{i\in I} \subset M$ be sub-modules of a left R-module M. Then the following are equivalent.*

1. *$\sum_{i\in I} N_i$ is direct and $\sum_{i\in I} N_i = M$.*
2. *For each $i \in I$, $N_i \cap \left(\sum_{j\neq i} N_j\right) = \{0\}$, and $\sum_{i\in I} N_i = M$.*
3. *Every $\underline{m} \in M$ has a unique representation as a finite sum of elements, each from a different N_i.*

Proof. First, suppose that (1) holds, and let $\varphi : \bigoplus_{i\in I} N_i \to M$ be the standard homomorphism. Pick $i \in I$, and suppose there exists $\underline{m_i} \in N_i$ and $\alpha_j \in R, \underline{m_j} \in N_j$ such that

$$\underline{m_i} = \sum_{j\neq i} \alpha_j \underline{m_j}$$

Defining $-\alpha_i = 1$, this implies in turn that

$$\varphi((\alpha_j \underline{m_j})_{j\in I}) = \underline{0}$$

But φ is an isomorphism, so it follows that $\underline{m_i} = \underline{0}$ and (2) holds.

Now, suppose that (2) holds, and that there exists two representations of some $\underline{m} \in M$ in the form of (3), say

$$\underline{m} = \sum_{i\in I} \underline{m_i} = \sum_{i\in I} \underline{x_i}$$

where $\underline{m_i}, \underline{x_i} \in N_i$. Without loss of generality, pick out some $i \in I$ such that $\underline{m_i} \neq \underline{x_i}$. Then we get

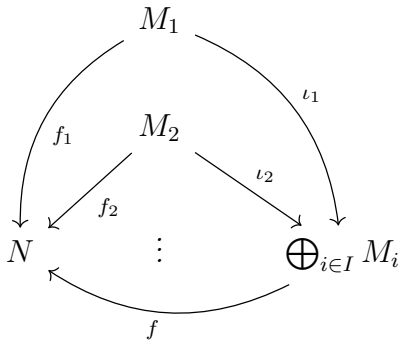
$$\underline{m_i} - \underline{x_i} \in N_i \backslash \{\underline{0}\}, \underline{m_i} - \underline{x_i} = \sum_{j\neq i} (\underline{x_j} - \underline{m_j}) \in \sum_{j\neq i} N_j$$

violating (2). Thus, (3) must hold.

Finally, suppose that (3) holds. That $\sum_{i\in I} N_i = M$ is immediate from this, so let $\varphi : \bigoplus_{i\in I} N_i \to M$ be the standard homomorphism. If $\ker(\varphi)$ is non-trivial, then $\underline{0}$ has two distinct representations in the form of (3). Thus, $\ker(\varphi)$ is trivial, making φ and isomorphism and (1) satisfied. □

The direct sum can also be characterized by the following universal property.

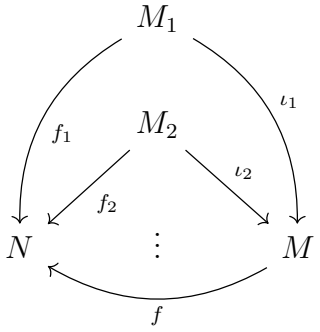
Theorem 4.3.5. *Let $\{M_i\}_{i\in I}$ be a collection of left R-modules with standard inclusions $\iota_i \in \mathrm{Hom}_R(M_i, \bigoplus_{i\in I} M_i)$, N some other R-module and $f_i \in \mathrm{Hom}_R(M_i, N)$ be linear maps. Then there exists a unique $f \in \mathrm{Hom}_R(\bigoplus_{i\in I} M_i, N)$ such that $f \circ \iota_i = f_i$. That is, the following diagram commutes.*



Proof. The condition $f \circ \iota_i = f_i$ forces the value of f on each $\iota_i(M_i)$, so since $\sum_{i\in I} \iota_i(M_i) = \bigoplus_{i\in M}$ this homomorphism is unique if it is well-define. To check that it is well-defined by the conditions $f \circ \iota_i = f_i$, we just need to check that it is single-valued, which is given by condition (2) in Theorem 4.3.4. □

In fact, we could have used this to define a direct sum. Indeed, starting at that point.

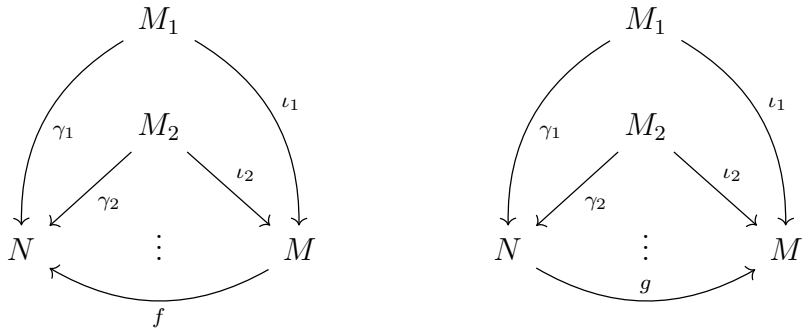
Definition 4.3.6. Let $\{M_i\}_{i\in I}$ be a collection of left R-modules. A direct sum of these modules is a left R-module M and collection of injective homomorphisms $\iota_i \in \mathrm{Hom}_R(M_i, M)$ satisfyping the following property : If N is some other R-module and $f_i \in \mathrm{Hom}_R(M_i, N)$ linear maps, then there exists a unique $f \in \mathrm{Hom}_R(M, N)$ such that $f \circ \iota_i = f_i$. That is, the following diagram commutes.



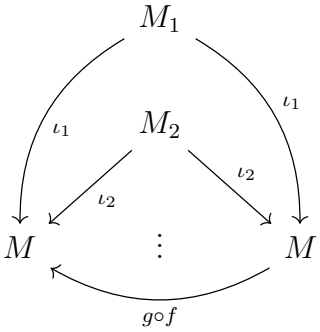
We can derive the following result.

Proposition 4.3.7. *Any pair of direct sums of a collection $\{M_i\}_{i\in I}$ of direct modules have a unique isomorphism between them satisfying the defining universal property.*

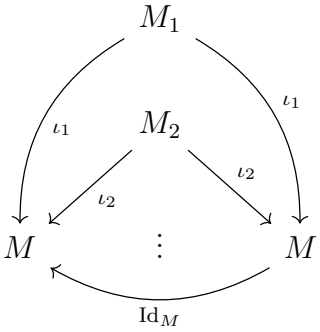
Proof. Suppose $(\{\iota_i\}_{i\in I}, M)$ and $(\{\gamma_i\}_{i\in I}, N)$ are two direct sums. Then there exists a unique homomorphism $f \in \mathrm{Hom}_R(M, N)$ and a unique homomorphism $g \in \mathrm{Hom}_R(N, M)$ such that the following two diagrams commute.



It thus follows that $(g \circ f) \circ \iota_i = g \circ (f \circ \iota_i) = g \circ \gamma_i = \iota_i$. Thus, the following diagram commutes.



But of course, the following diagram also commutes

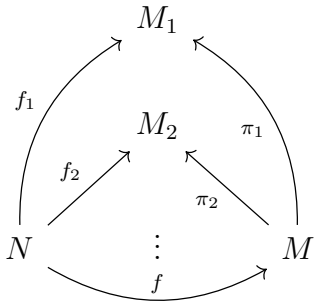


so by the uniqueness property of the direct sum we conclude that $g \circ f = \mathrm{Id}_M$. A similar argument shows that $f \circ g = \mathrm{Id}_N$, so f is the desired isomorphism. By the uniqueness property of the direct sum, f is the only such isomorphism. □

Essentially what we're saying here is that given any pair of direct sums, we can find a unique *compatible* isomorphism between them. Hence, there is in a sense only one direct sum construction.

We'll next do our first example of what's called *dualizing*. The premise is quite simple, what would happen if we reversed all the arrows in definition 4.3.6?

Definition 4.3.8. Let $\{M_i\}_{i\in I}$ be a collection of left R-modules. A direct product of these modules is a left R-module M and collection of surjective homomorphisms $\pi_i \in \mathrm{Hom}_R(M, M_i)$ satisfyping the following property : If N is some other R-module and $f_i \in \mathrm{Hom}_R(N, M_i)$ linear maps, then there exists a unique $f \in \mathrm{Hom}_R(N, M)$ such that $\pi_i \circ f = f_i$. That is, the following diagram commutes.



The first thing to check is that something actually satisfies this definition. What turns out to work is essentially duplicating the original construction of the direct sum, but allowing all sequences instead of just those with finitely many non-zero elements. It will be left to the reader to check that this is the desired direct product, and that direct products, like direct sums, are unique up to a unique compatible isomorphism. In note of this, we denote the direct sum (rather confusingly) by $\prod_{i\in I} M_i$.

There is something very important to notice here. Our definitions of the direct product and sum depend only on *the properties of homomorphisms*. So there's nothing to stop us from taking these definitions and porting them over to rings or groups. Indeed, the direct products and sums that have shown up in previous chapters could be defined by the same universal property! We'll look into this more in Part III.

Note. In the case of left R-modules, *finite* direct sums and products are the same, but *infinite* ones are different. This does not necessarily carry over to other algebraic objects.

4.4 Matrices Over Principle Idea Domains

From now on, we assume all our rings are PIDs and hence commutative. There will no longer be a distinction between left and right modules as a result, and we simply call both modules.

TODO

4.5 Structure Theorem

Chapter 5

Free Commutative Modules

Part III

The Abstract View

Chapter 6

Universal Algebras

6.1 Universal Algebras

Let us start with a construction which will seem rather abstract now, but becomes powerful in practice. Much of this section is based of the results in [BS12].

Definition 6.1.1. A universal algebra $\mathcal{A} = \langle \mathcal{U}, \mathcal{F} \rangle$ is a set (called the universe) $\mathcal{U}$ along with a family of operations $\mathcal{F}$ from $\mathcal{U}^n$ to $\mathcal{U}$ of finite arity.

Example 6.1.1. A group is a universal algebra $\langle G, \cdot, ^{-1}, 1 \rangle$ with 2-ary, 1-ary, and 0-ary operations respectively satisfying the following axioms, for all $x, y, z \in G$

1. $x \cdot (y \cdot z) = (x \cdot y) \cdot z$
2. $x \cdot 1 = 1 \cdot x = x$
3. $x \cdot x^{-1} = x^{-1} \cdot x = 1$

A group is called commutative (or Abelian) if the additional axiom $x \cdot y = y \cdot x$ is satisfied.

Example 6.1.2. A monoid is a group without the 1-ary operation.

Example 6.1.3. A ring is a universal algebra $\langle R, \cdot, 1, +, -, 0 \rangle$ such that

1. $\langle R, +, -, 0 \rangle$ is a commutative group
2. $\langle R, \cdot, 1 \rangle$ is a monoid
3. $\forall x, y, z \in R$,

$$x \cdot (y + z) = x \cdot y + x \cdot z \qquad (x + y) \cdot z = x \cdot z + y \cdot z$$

Example 6.1.4. A left monoid over a ring R is a universal algebra $\langle M, +, -, 0, \{f_r\}_{r \in R} \rangle$ such that

1. $\langle M, +, -, 0 \rangle$ is a commutative group

2. $\forall x, y \in M, r \in R,\ f_r(x+y) = f_r(x) + f_r(y)$

3. $\forall x \in M, r, s \in R,\ f_{r+s}(x) = f_r(x) + f_s(x)$ and $f_{r\cdot s}(x) = f_r(f_s(x))$

Definition 6.1.2. A sub-universe of a universal algebra $\mathcal{A} = \langle \mathcal{U}, \mathcal{F} \rangle$ is a subset $\mathcal{S} \subseteq \mathcal{U}$ such that $\langle \mathcal{S}, \mathcal{F}' \rangle$ is a universal algebra, where $\mathcal{F}' = \{f_{|\mathcal{S}} \mid f \in \mathcal{F}\}$. $\langle \mathcal{S}, \mathcal{F}' \rangle$ is a sub-algebra of $\mathcal{A}$.

Definition 6.1.3. A congruence $\equiv$ on a universal algebra $\mathcal{A} = \langle \mathcal{U}, \mathcal{F} \rangle$ is an equivalence relation on $\mathcal{U}$ such that for any n-ary operation $f \in \mathcal{F}$ and $a_i, a_i' \in \mathcal{U}$

$$a_i \equiv a_i' \Rightarrow f(a_1, \ldots, a_n) \equiv f(a_1', \ldots, a_n')$$

The notion of congruence allows us to define a particular type of universal algebra, which we call the quotient algebra.

Theorem 6.1.4. *Let $\mathcal{A} = \langle A, \mathcal{F} \rangle$ be a universal algebra and $\equiv$ a congruence on $\mathcal{A}$. Then ${}^{\mathcal{A}}/_{\equiv} = \langle {}^{A}/_{\equiv}, \mathcal{F} \rangle$ is a well-defined universal algebra, where an n-ary $f \in \mathcal{F}$ acts by, for any $[a_1], \ldots, [a_n] \in {}^{A}/_{\equiv}$*

$$f([a_1], \ldots, [a_n]) = [f(a_1, \ldots, a_n)]$$

Proof. It suffices to show that the action of each n-ary $f \in \mathcal{F}$ is well-defined. Indeed, suppose that $[a_i] = [b_i] \in {}^{A}/_{\equiv}$. Then since $\equiv$ is a congruence and $a_i \equiv b_i$,

$$f(a_1, \ldots, a_n) \equiv f(b_1, \ldots, b_n) \Rightarrow [f(a_1, \ldots, a_n)] = [f(b_1, \ldots, b_n)]$$

as required. □

This ${}^{\mathcal{A}}/_{\equiv}$ is called the *quotient* of $\mathcal{A}$ by $\equiv$. Finally, let $A' \subseteq A$ be an arbitrary subset. We define the *universal algebra generated by* A', denoted $\langle A' \rangle$ to be the intersection of all universes B containing A' such that $\langle B, \mathcal{F} \rangle$ is a universal algebra. This is well-defined as long as the intersection of universes with operations $\mathcal{F}$ is a universe, the verification of which is left to the reader.

At this point, we begin to play a bit fast and loose with definitions to avoid some more abstract concepts. We'll say that two universal algebras are of the same *type* if they have a corresponding set of n-ary operations, that is their n-ary operations can be put in bijective correspondence for each $n \in \mathbb{N} \cup \{0\}$. The most basic example of this would be to note that any sub-algebra is of the same type as its parent algebra. For a more precise notion than this, see [BS12].

Definition 6.1.5. Let $\mathcal{A} = \langle A, \mathcal{F} \rangle, \mathcal{B} = \langle B, \mathcal{F} \rangle$ be universal algebras of the same type. A map $\varphi : A \to B$ is called a homomorphism if, for any n-ary operation $f \in \mathcal{F}$ and $a_i \in A$

$$f(\varphi(a_1), \ldots, \varphi(a_n)) = \varphi(f(a_1, \ldots, a_n))$$

Note. There's a bit of an abuse of notation here, as f is technically two different operations on A and B. We regard it as acting on both by the pairing between n-ary operations of the two universal algebras. This notation will be used for the remainder of this section.

Note. A nice property of homomorphisms, the proof of which is left to the reader, is that the composition of two homomorphisms is a homomorphism.

An injective homomorphism is called a *monomorphism*, a surjective homomorphism an *epimorphism*, and a bijective homomorphism an *isomorphism*. An isomorphism from a universe to itself is called an *automorphism*, and two universal algebras are called *isomorphic* if there exists an isomorphism between them, which is denoted by the symbol $\cong$. Homomorphism are of vital importance of algebra, as their "preservation" of operations allows us to determine when two universal algebras are essentially identical. In order to make this more precise, we need the following two lemmas.

Lemma 6.1.6. *Suppose $\mathcal{A}, \mathcal{B}$ are universal algebras of the same type, and $\varphi : A \to B$ a homomorphism. Then $\varphi(\mathcal{A}) = \langle \varphi(A), \mathcal{F} \rangle$ is a sub-algebra of $\mathcal{B}$, and $\varphi^{-1}(\mathcal{B}) = \langle \varphi^{-1}(B), \mathcal{F} \rangle$ is a sub-algebra of $\mathcal{A}$.*

Proof. Let $f \in \mathcal{F}$ be an n-ary operation. If $b_1, \ldots, b_n \in \varphi(A)$, then $\exists a_i \in A$ such that $\varphi(a_i) = b_i$. Thus,

$$f(b_1, \ldots, b_n) = f(\varphi(a_1), \ldots, \varphi(a_n)) = \varphi(f(a_1, \ldots, a_n)) \in \varphi(A)$$

so $\varphi(\mathcal{B})$ is a sub-algebra, as claimed. The proof of the second statement is essentially identical. □

Lemma 6.1.7. *Suppose $\mathcal{A}, \mathcal{B}$ are universal algebras of the same type, and $\varphi : A \to B$ a homomorphism. Then the equivalence relation $\equiv$ on A given by $a_1 \equiv a_2$ if $\varphi(a_1) = \varphi(a_2)$ is a congruence.*

Proof. Suppose $f \in \mathcal{F}$ is an n-ary operation, and $a_i \equiv a_i' \in A$. Then

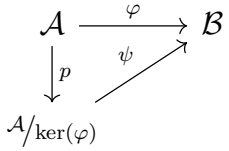
$$\varphi(f(a_1, \ldots, a_n)) = f(\varphi(a_1), \ldots, \varphi(a_n)) = f(\varphi(a_1'), \ldots, \varphi(a_n')) = \varphi(f(a_1', \ldots, a_n'))$$

so $f(a_1, \ldots, a_n) \equiv f(a_1', \ldots, a_n')$, as required. □

In the case of the above lemma, we denote that $\mathcal{A}/\equiv = \mathcal{A}/\ker(\varphi)$, and $A/\equiv = A/\ker(\varphi)$.

With these, we finally reach the fundamental theorems of homomorphisms.

Theorem 6.1.8 (The First Fundamental Theorem of Homomorphisms)**.** *Suppose $\mathcal{A}, \mathcal{B}$ are universal algebras of the same type, and $\varphi : A \to B$ a homomorphism. Then the projection map $p : A \to A/\ker(A)$ is a homomorphism, and there exists an isomorphism $\psi : \mathcal{A}/\ker(\varphi) \to \varphi(\mathcal{A})$ such that the following diagram commutes*



In particular, $\mathcal{A}/\ker(\varphi) \cong \varphi(\mathcal{B})$.

Proof. The first statement follows from the projection map onto equivalence classes for a congruence always being a homomorphism, which is easily verified and left to the reader. We define $\psi : {}^{A}\!/_{\ker(A)} \to \varphi(A)$ by, for any $[a] \in {}^{A}\!/_{\ker(A)}$, $\psi([a]) = \psi(a)$. It remains to show that this is in fact well-defined, a homomorphism, bijective, and satisfies the commutative property. For the first, we note that if $[a] = [a'] \in {}^{A}\!/_{\ker(A)}$, then $\varphi(a) = \varphi(a')$, making ψ well-defined. Pick any n-ary $f \in \mathcal{F}$, and $[a_1], \ldots, [a_n] \in {}^{A}\!/_{\ker(A)}$. Then

$$f(\psi([a_1]), \ldots, \psi([a_n])) = f(\varphi(a_1), \ldots, \varphi(a_n)) = \varphi(f(a_1, \ldots, a_n)) = \varphi(f([a_1], \ldots, [a_n]))$$

so ψ is a homomorphism. Pick any $b \in \varphi(B)$. Then $\exists a \in A$ such that $\varphi(a) = b$, so $\psi([a]) = b$ and hence ψ is surjective. Suppose $\psi([a_1]) = \psi([a_2])$, where $[a_1], [a_2] \in {}^{A}\!/_{\ker(A)}$ Then $\varphi(a_1) = \varphi(a_2) \Rightarrow [a_1] = [a_2]$, so ψ is injective and hence bijective. Finally, we check the commutativity property. Let $a \in A$, then by definition

$$\varphi(a) = \psi([a]) = (\psi \circ p)(a)$$

as claimed. □

For the next two isomorphism theorems, we need the notion of a sub-congruence.

Definition 6.1.9. Let $\mathcal{A}$ be a universal algebra, and $\equiv$ a congruence on $\mathcal{A}$. A congruence $\sim$ on $\mathcal{A}$ is called a sub-congruence of $\equiv$ if $a \equiv a' \Rightarrow a \sim a'$, for all $a, a' \in A$.

Lemma 6.1.10. *Let $\equiv$ be a congruence on a universal algebra $\mathcal{A}$, and $\sim$ a sub-congruence of $\equiv$. Let ${}^{\equiv}\!/_{\sim}$ be the equivalence relation on ${}^{\mathcal{A}}\!/_{\sim}$ defined by*

$$[a]_{\sim} {}^{\equiv}\!/_{\sim} [a']_{\sim} \iff a \equiv a'$$

This is a well-defined equivalence relation, and a congruence on ${}^{\mathcal{A}}\!/_{\sim}$

Proof. We start by show that it is an equivalence relation. Reflexivity and symmetry are clear, so it suffices to show transitivity. Suppose that $[a], [a'], [a''] \in {}^{\mathcal{A}}\!/_{\sim}$ are such that $[a] {}^{\equiv}\!/_{\sim} [a']$ and $[a'] {}^{\equiv}\!/_{\sim} [a'']$. Then $a \equiv a'$ and $a' \equiv a''$, so $a \equiv a''$ and hence $[a] {}^{\equiv}\!/_{\sim} [a'']$, as required. Now, let $f \in \mathcal{F}$ be an n-ary operation. Suppose $[a_i] {}^{\equiv}\!/_{\sim} [b_i] \in {}^{\mathcal{A}}\!/_{\sim}$. Then $a_i \equiv b_i$, so

$$f([a_1], \ldots, [a_n]) = [f(a_1, \ldots, a_n)] {}^{\equiv}\!/_{\sim} [f(b_1, \ldots, b_n)] = f([b_1], \ldots, [b_n])$$

making ${}^{\equiv}\!/_{\sim}$ a congruence as claimed. □

Theorem 6.1.11 (The Second Fundamental Theorem of Homomorphisms)**.** *Let $\equiv$ be a congruence on a universal algebra $\mathcal{A}$, and $\sim$ a sub-congruence of $\equiv$. Then there exists an isomorphism $\varphi : {}^{({}^{\mathcal{A}}\!/_{\sim})}\!/_{({}^{\equiv}\!/_{\sim})} \to {}^{\mathcal{A}}\!/_{\equiv}$ such that the following diagram commutes.*

$$\begin{array}{ccc} \mathcal{A} & \longrightarrow & {}^{\mathcal{A}}\!/_{\equiv} \\ \downarrow & & \uparrow{\scriptstyle \varphi} \\ {}^{\mathcal{A}}\!/_{\sim} & \longrightarrow & {}^{({}^{\mathcal{A}}\!/_{\sim})}\!/_{({}^{\equiv}\!/_{\sim})} \end{array}$$

where all the unlabelled arrows are the natural projection maps. In particular, $\mathcal{A}/\equiv \cong (\mathcal{A}/\sim)/(\equiv/\sim)$.

Proof. We define φ by the rule, for any $a \in A$, $\varphi([[a]_\sim]_{\equiv/\sim}) = [a]_\equiv$. We first show that it is a homomorphism. Pick any n-ary $f \in \mathcal{F}$, $a_i \in A$. Then

$$\begin{aligned} f(\varphi([[a_1]_\sim]_{\equiv/\sim}), \ldots, \varphi([[a_n]_\sim]_{\equiv/\sim})) &= f([a_1]_\equiv, \ldots, [a_n]_\equiv) = [f(a_1, \ldots, a_n)]_\equiv \\ &= \varphi([[f(a_1, \ldots, a_n)]_\sim]_{\equiv/\sim}) = \varphi(f([[a_1]_\sim]_{\equiv/\sim}, \ldots, [[a_n]_\sim]_{\equiv/\sim})) \end{aligned}$$

as required. Next, we show that it's injective. Suppose that $\varphi([[a]_\sim]_{\equiv/\sim}) = \varphi([[a']_\sim]_{\equiv/\sim})$ for some $a, a' \in A$. Then $[a]_\equiv = [a']_\equiv \Rightarrow a \equiv a' \Rightarrow [a]_\sim \overset{\equiv}{/\sim} [a']_\sim \Rightarrow [[a]_\sim]_{\equiv/\sim} = [[a']_\sim]_{\equiv/\sim}$ as required. Surjectivity is clear, so φ is an isomorphism. Finally, we show that the commutativity property is satisfied. Pick any $a \in A$. Then

$$(\varphi \circ p_{\equiv/\sim} \circ p_\sim)(a) = (\varphi \circ p_{\equiv/\sim})([a]_\sim) = \varphi([[a]_\sim]_{\equiv/\sim}) = [a]_\equiv = p_\equiv(a)$$

as required. □

Theorem 6.1.12 (The Third Fundamental Theorem of Homomorphisms)**.** *Suppose* $\mathcal{B}$ *is a subalgebra of* $\mathcal{A}$*, and* $\equiv$ *is a congruence on* $\mathcal{A}$*. Set* $B^\equiv = \{a \in A \mid B \cap [a]_\equiv \neq \emptyset\}$*. Then*

1. $\equiv_{|_B}$ *is a congruence*
2. $\mathcal{B}^\equiv$ *is a subalgebra of* $\mathcal{A}$
3. $\mathcal{B}/\equiv_{|_B} \cong \mathcal{B}^\equiv/\equiv_{|_{B^\equiv}}$

Proof. (1) is fairly clear, and its proof will be left to the reader. For (2), pick any n-ary $f \in \mathcal{F}$ and $a_1, \ldots, a_n \in B^\equiv$. It suffices to show that $f(a_1, \ldots, a_n) \in B^\equiv$. Indeed, we can note that for each a_i, there exists some $b_i \in B$ such that $a_i \equiv b_i$, and hence

$$[f(a_1, \ldots, a_n)]_\equiv = f([a_1]_\equiv, \ldots, [a_n]_\equiv) = f([b_1]_\equiv, \ldots, [b_n]_\equiv) = [f(b_1, \ldots, b_n)]_\equiv$$

Since $\mathcal{B}$ is a subalgebra, $f(b_1, \ldots, b_n) \in B$, so it follows that $[f(a_1, \ldots, a_n)]_\equiv \cap B \neq \emptyset$ and hence $f(a_1, \ldots, a_n) \in B^\equiv$ as required. It is also good to note that this further shows that $\mathcal{B}$ is a subalgebra of $\mathcal{B}^\equiv$. For (3), we first note that (1) implies that $\equiv_{|_{B^\equiv}}$ is a congruence, so this statement is well-defined. We define our isomorphism $\varphi : B/\equiv_{|_B} \to B^\equiv/\equiv_{|_{B^\equiv}}$ by, for any $b \in B$, $\varphi([b]_{\equiv_{|_B}}) = [b]_{\equiv_{|_{B^\equiv}}}$. We first show that this is a homomorphism. Pick any n-ary $f \in \mathcal{F}$ and $b_1, \ldots, b_n \in B$. Then

$$\begin{aligned} f(\varphi([b_1]_{\equiv_{|_B}}), \ldots, \varphi([b_n]_{\equiv_{|_B}})) &= f([b_1]_{\equiv_{|_{B^\equiv}}}, \ldots, [b_n]_{\equiv_{|_{B^\equiv}}}) = [f(b_1, \ldots, b_n)]_{\equiv_{|_{B^\equiv}}} \\ &= \varphi([f(b_1, \ldots, b_n)]_{\equiv_{|_B}}) = \varphi(f([b_1]_{\equiv_{|_B}}, \ldots, [b_n]_{\equiv_{|_B}})) \end{aligned}$$

as required. Surjectivity is clear. To show that φ is injective, suppose that $b, b' \in B$ are such that $\varphi([b]_{B_{|_\equiv}}) = \varphi([b']_{B_{|_\equiv}})$. Then $[b]_{\equiv_{|_{B^\equiv}}} = [b']_{\equiv_{|_{B^\equiv}}} \Rightarrow [b]_{B_{|_\equiv}} = [b']_{B_{|_\equiv}}$ as required. □

These theorems will reappear in different guises for groups in section 2.4, rings in section 3.4, and modules in **??**.

6.2 Direct Products

This section is based on results from [BS12] and [Sil23].

In this section, we cover the one of the most common constructions on spaces in algebra, the direct product. Before doing this, we have a quick bit of notation : we denote the collection of all homomorphisms between universal algebras $\mathcal{A}, \mathcal{B}$ by $\mathrm{Hom}(\mathcal{A}, \mathcal{B})$.

Definition 6.2.1. Let $\{\mathcal{A}_i\}_{i\in I}$ be an indexed family of universal algebras of the same type. We call another universal algebra of the same type $\mathcal{B}$ a direct product of $\{\mathcal{A}_i\}_{i\in I}$ if there exist homomorphisms $\pi_i : B \to A_i$ such that for any other universal algebra of the same type $\mathcal{C}$ and homomorphisms $\{\psi_i : C \to A_i\}_{i\in I}$, there exists a unique homomorphism $\varphi \in \mathrm{Hom}(\mathcal{C}, \mathcal{B})$ such that $\pi_i \circ \varphi = \psi_i$ for all $i \in I$.

This of course is quite an abstract definition. Luckily, for universal algebras we have the following result to make things more tangible.

Theorem 6.2.2. *For any indexed family $\{\mathcal{A}_i\}_{i\in I}$ of universal algebras of the same type, there exists a direct product unique up to a unique isomorphism.*

Proof. We start with uniqueness. Suppose $\mathcal{B}, \mathcal{C}$ are direct products of $\{\mathcal{A}_i\}_{i\in I}$. Then there exists a unique homomorphism $\varphi \in \mathrm{Hom}(\mathcal{C}, \mathcal{B})$ such that $\pi_{i,B} \circ \varphi = \pi_{i,C}$, and there exists a unique homomorphism $\zeta \in \mathrm{Hom}(\mathcal{B}, \mathcal{C})$ such that $\pi_{i,C} \circ \zeta = \pi_{i,B}$. It follows that $\pi_{i,B} \circ \varphi \circ \zeta = \pi_{i,B}$, so $\varphi \circ \zeta = \mathrm{Id}_{\mathcal{C}}$. Similarly, $\zeta \circ \varphi = \mathrm{Id}_{\mathcal{B}}$, so φ is an isomorphism. By the uniqueness of φ, ζ, it is the unique isomorphism between $\mathcal{C}, \mathcal{B}$, as claimed.

Next, we show existence. We define the universal algebra $\prod_{i\in I} \mathcal{A}_i$ as having the universe $\prod_{i\in I} A_i$, and operations defined by, for any $f \in \mathcal{F}$

$$f((a_{i1})_{i\in I}, \ldots, (a_{in})_{i\in I}) = (f(a_{i1}, \ldots, a_{in}))_{i\in I}$$

where $a_{ij} \in \mathcal{A}_i$. That this is a universal algebra is clear, and since projection maps between universal algebras are homomorphisms, we have our required homomorphisms $\pi_j \in \mathrm{Hom}(\prod_{i\in I} \mathcal{A}_i, \mathcal{A}_j)$. It suffices then to show that these satisfy the desired property. Suppose $\mathcal{C}$ is another universal algebra of the same type, and $\psi_j \in \mathrm{Hom}(\mathcal{C}, \mathcal{A}_j)$ homomorphisms. If we try to define $\varphi : C \to B$ by $\pi_i \circ \varphi = \psi_i$, we can note that this requires $\varphi(c) = \prod_{i\in I} \psi_i(c)$ for any $c \in C$, giving the uniqueness of φ. It suffices then to prove that φ is a homomorphism. Pick any $f \in \mathcal{F}$ and $c_i \in C$. Then

$$\begin{aligned} f(\varphi(c_1), \ldots, \varphi(c_n)) &= f\Big(\prod_{i\in I} \psi_i(c_1), \ldots, \prod_{i\in I} \psi_i(c_n)\Big) = \prod_{i\in I} f(\psi_i(c_1), \ldots, \psi_i(c_n)) \\ &= \prod_{i\in I} \psi_i(f(c_1, \ldots, c_n)) = \varphi(f(c_1, \ldots, c_n)) \end{aligned}$$

as required. □

Note. It is not too difficult to show, using the above theorem, that the direct product is associative and commutative up to isomorphism.

There's another important construction to mention here, called the *direct sum*.

Definition 6.2.3. Let $\{\mathcal{A}_i\}_{i\in I}$ be an indexed family of universal algebras of the same type. We call another universal algebra of the same type $\mathcal{B}$ a direct sum of $\{\mathcal{A}_i\}_{i\in I}$ if there exist homomorphisms $\iota_i : A_i \to B$ such that any other universal algebra of the same type $\mathcal{C}$ and homomorphisms $\{\psi_i : A_i \to C\}_{i\in I}$, there exists a unique homomorphism $\varphi \in \mathrm{Hom}(\mathcal{B}, \mathcal{C})$ such that $\varphi \circ \iota_i = \psi_i$ for all $i \in I$.

This is essentially just the definition of the direct product with the arrows reversed. Similarly to the direct product, one could show that the direct sum is unique up to unique isomorphism when it exists. Existence is a much trickier proposition, as the structure of a direct sum can get exceedingly complicated[1]. As such, we will treat direct sums in the subsequent chapters on a case-by-case basis. A more unified description of direct sums and when they exist is possible though category theory, the interested reader is directed to [Lan10] and [Lan05].

[1] I suspect that it's not guaranteed for all universal algebras, but I have not yet thought of a counterexample

Chapter 7

Categories

7.1 Basic Definitions

7.2 Dualizing

7.3 Universal Properties and Examples

7.4 The Yoneda Lemma

Part IV

Advanced Algebra

Chapter 8

Field Extensions and Galois Theory

Chapter 9

Commutative Algebra

Chapter 10

Homology

Bibliography

[Bad10] B. Badzioch, Math 619 : Abstract algebra i lecture notes, week 13, http://www.math.buffalo.edu/~badzioch/MTH619/Lecture_Notes.html, 2010.

[BS12] S. Burris and H. Sankappanavar, *A Course in Universal Algebra*, mellenium ed., 2012.

[Gub21] N. Gubareni, *Introduction to Modern Algebra and Its Applications*, 1. ed., CRC Press, 2021.

[Jac09] N. Jacobson, *Basic Algebra I*, second ed., Dover Publications, Mineola, NY, 2009.

[Lan10] S. M. Lane, *Categories for the Working Mathematician*, second ed., Springer-Verlag, New York, NY, 2010.

[Lan05] S. Lang, *Algebra*, third ed., Springer-Verlag, New York, NY, 2005.

[Rot07] S. Rotman, *Advanced Linear Algebra*, third ed., Springer-Verlag, New York, NY, 2007.

[Sil23] L. Silberman, Math 412: Advanced linear algebra lecture notes, https://personal.math.ubc.ca/~lior/teaching/2223/412_W23/, 2023.