Abstract Algebra Review Notes

Will Huang
UW-Madison

Updated July 11, 2021

Contents

Part I
Introduction

These notes were compiled from the following classes during the following semesters to be used as a reference for future courses.

- Math 541: Modern Algebra (Spring 2019) - A general undergraduate introduction to groups and rings along with examples and applications.
- Math 542: Modern Algebra (Fall 2019) - Continuation of 542 covering modules and fields along with a brief introduction to Galois Theory.
- Math 741: Abstract Algebra (Fall 2020) - Intended to cover topics on the algebra qualifying exam, includes a more in depth treatment of groups and rings along with some representation and category theory.
- Math 742: Abstract Algebra (Spring 2021) - Graduate level treatment of commutative algebra and Galois Theory.

Graduate and undergraduate level content are intermingled for the purpose of keeping topics together. If that annoys you then sucks to suck.

The textbooks referenced/used in the courses are below

- Dummit and Foote - Abstract Algebra
- MacLane - Categories for the Working Mathematician
- Fulton - A First Course in Representation Theory
- Atiyah and McDonald - Commutative Algebra
- Milne - Fields and Galois Theory [1]

1 Properties of Integers

1.1 Residue Classes of Integers

Before we introduce any actual algebra, let's first examine something that we are all familiar with: the integers. We can think of this as a case study or motivating example.

Definition. *Given any two integers, the greatest common divisor (GCD) is the greatest integer dividing both.*

Let $d = gcd(a, b)$ where $a, b \in \mathbb{Z}$. We can decompose any GCD in terms of its operands. That is, $\exists x, y \in \mathbb{Z}$ such that $d = ax + by$, the decomposition can be found using the Euclidean algorithm demonstrated below.

[1] available for free here: https://www.jmilne.org/math/CourseNotes/ft.html

Example. Decompose 4 in terms of 3084 and 1424 (gcd(3084,1424) = 4)

$$\begin{aligned} 3084 = 2 \cdot 1424 + 236 \quad &\rightarrow \quad 236 = 3084 - 2 \cdot 1424 \\ 1424 = 6 \cdot 236 + 8 \quad &\rightarrow \quad 8 = 1424 - 6 \cdot 236 \\ 236 = 29 \cdot 8 + 4 \quad &\rightarrow \quad 4 = 236 - 29 \cdot 8 \end{aligned}$$

Working backwards will give us the required decomposition

$$\begin{aligned} 4 &= 236 - 29 \cdot 8 \\ &= 236 - 29(1424 - 6 \cdot 236) \\ &= 175 \cdot 236 - 29 \cdot 1424 \\ &= 175(3084 - 2 \cdot 1424) - 2 \cdot 1424 \\ &= 175 \cdot 3084 - 379 \cdot 1424 \end{aligned}$$

We give a formal definition to the concept of remainders when dividing integers by defining the modulo operator. In fact, we find that this defines a set of equivalence classes.

Definition. *For any two integers a, b, we say a is congruent to $b \mod n$ if $n|(b-a)$. In mathematical terms*

$$a \equiv b \mod n \rightarrow \exists k \in \mathbb{Z} \text{ s.t. } a = b + kn$$

Proposition 1.1. The operation $\equiv \mod n$ defines an equivalence relation

Proof. To show an operation defines an equivalence relation, we must show reflexivity, symmetry and transitivity. If $a \equiv a \mod n$, then $a - a = 0 = 0 \cdot n$, thus an integer is in its own equivalence class and modulo is reflexive.
Suppose $a \equiv b \mod n$, then $b - a = kn \rightarrow a - b = -kn \rightarrow b \equiv a \mod n$. Since the integers are closed under negatives, modulo is symmetric.
Suppose $a \equiv b \mod n$, $b \equiv c \mod n$. Then $b - a = k_1 n$, $c - b = k_2 n \rightarrow c - a = (k_1 + k_2)n \rightarrow a \equiv c \mod n$ and so modulo is transitive. Since all three properties hold, we conclude that modulo defines an equivalence relation. □

The equivalence classes of the modulo operator are called residue classes and denoted

$$\overline{a} = \{b \in \mathbb{Z} \mid b = a \mod n\}$$

Since equivalence classes partition a set, we define the integers mod n as follows

$$\mathbb{Z}/n\mathbb{Z} = \{\overline{a} \mid a \in \mathbb{Z}\} = \left\{\overline{0}, \overline{1}, \ldots, \overline{n-1}\right\}$$

Two residue classes are equal if they are congruent mod n. That is

$$\overline{a} = \overline{b} \rightarrow a \equiv b \mod n$$

Furthermore, we can define operations on these residue classes.

Proposition 1.2. Suppose $a_1 \equiv a_2 \mod n$ and $b_1 \equiv b_2 \mod n$, then $a_1 + b_1 \equiv (a_2 + b_2) \mod n$ and $a_1 b_1 \equiv a_2 b_2 \mod n$. Thus for any two equivalence classes $\overline{a}, \overline{b} \in \mathbb{Z}/n\mathbb{Z}$, we can define addition and multiplication linearly

$$\overline{a} + \overline{b} = \overline{a + b} \qquad \overline{a} \cdot \overline{b} = \overline{ab}$$

It's clear the set of integers mod n is closed and invertible under addition, that is every element has another element that sum to zero. Unfortunately, it is slightly more difficult to achieve this multiplicatively where the identity is 1. Fortunately, by discarding a few elements we can define the multiplicative group of integers mod n.

$$(\mathbb{Z}/n\mathbb{Z})^\times = \{\overline{a} \in \mathbb{Z}/n\mathbb{Z} \mid \exists c \in \mathbb{Z},\, \overline{ac} = 1\}$$

Example. $(\mathbb{Z}/12\mathbb{Z})^\times = \left\{\overline{1}, \overline{5}, \overline{7}, \overline{11}\right\}$

Proposition 1.3. The elements of $(\mathbb{Z}/n\mathbb{Z})^\times$ are simply the elements of the additive group that are coprime to n, that is $gcd(a, n) = 1 \to \overline{a} \in (\mathbb{Z}/n\mathbb{Z})^\times$

Proof. Suppose $gcd(a, n) = d > 1$, which means $\exists x, y \in \mathbb{Z}$ such that $a = dx, n = dy$. Where $gcd(x, y) = 1$. Thus $ay = xyb = xn \to ay \equiv 0 \mod n$. If we assume that there exists some integer c such that $ac \equiv 1 \mod n$, then

$$ac \equiv 1 \mod n \to acy \equiv y \mod n \to 0 \equiv y \mod n$$

which is a contradiction as $y < n$, therefore such an integer c cannot exist. Thus $gcd(a, n) \neq 1 \to \overline{a} \notin (\mathbb{Z}/n\mathbb{Z})^\times$
Now suppose $gcd(a, n) = 1$, then using the Euclidean Algorithm gives

$$1 = ax + by \to ax = 1 + n(-y) \to ax \equiv 1 \mod n$$

In other words, $gcd(a, n) = 1 \to \overline{a} \in (\mathbb{Z}/n\mathbb{Z})^\times$ as required. □

Part II
Group Theory

2 Introduction to Groups

2.1 Definition and Properties

Definition. *A group is a set G along with an operation $\cdot : G \times G \to G$ such that*

1. *$\cdot$ is associative: $(ab)c = a(bc)$*
2. *An identity element, denoted 1, exists: $1 \cdot a = a = a \cdot 1$*
3. *The set is closed under inverses, that is every element a in the group has an inverse b also in the group such that $ab = 1 = ba$*

If the group operation is commutative $ab = ba$ then the group is said to be abelian.

Example. The following are all groups

1. The set $\mathbb{Z}$ along with normal addition $+$ defines a group
2. The set of vectors $\mathbb{Z}^n$ with element-wise addition defines an abelian group
3. The set of non zero rational numbers $\mathbb{Q}^\times$ under multiplication is an abelian group
4. The set of $n \times n$ matrices with nonzero determinants forms a group under multiplication, this is called the general linear group and denoted $GL_n(\mathbb{Z})$. All matrices in this group will have ± 1 as a determinant.
5. The set of matrices with determinant 1, the special linear group, forms a group under matrix multiplication and is denoted $SL_n(\mathbb{Z})$

Proposition 2.1 (Properties of Groups)**.**
There are some useful group properties to note

1. The identity of a group is unique
2. The inverse of any group element is unique
3. "Cancellation" holds: $au = bu \to a = b$
4. The inverse of an inverse is itself: $(a^{-1})^{-1} = a$. Furthermore the inverse of a product reverses the order: $(ab)^{-1} = b^{-1}a^{-1}$ (shoes and socks)

Proof. We will prove these properties one at a time

1. Suppose $1 \neq 1'$ are unique inverses, then $1 = 1 \cdot 1' = 1' \cdot 1 = 1'$ which is clearly a contradiction, so the two must be equal.

2. Suppose a has two inverses b and c, then $b = b \cdot 1 = b(ac) = (bc)c = 1 \cdot c = c$, thus the two must be the same.

3. Suppose $au = bu$, then $auu^{-1} = buu^{-1} \to a \cdot 1 = b \cdot 1 \to a = b$. A similar argument shows that left cancellation also applies.

4. $a^{-1}a = aa^{-1} = 1 \to (a^{-1})^{-1} = a$. To show that products are reversed under inversion, $(ab)(b^{-1}a^{-1}) = a(bb^{-1})a^{-1} = a \cdot 1 \cdot a^{-1} = 1$.

Note that property 4 is due to the group action generally not being commutative, in an abelian group we can simply reverse the order of the product and so the property would not matter. □

Definition. *The order of an element $a \in G$ is the smallest integer n such that $a^n = 1$, this is denoted $|a| = n$*

We can also define the order of a group as the number of the elements it contains, this will also be denoted $|G|$

Example. We can show through repetitive multiplication that

$$\left| \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \right| = 4$$

There are a few noteworthy groups that are frequently used throughout algebra which we will examine these groups in depth here. We can represent the integers 1 to n with

$$[n] = \{1, 2, 3, ..., n\}$$

Definition. *A permutation is a bijective function $\sigma : [n] \to [n]$*

1. *The product of two permutations is also a permutation*

2. *Permutations are invertible*

In case the last two properties haven't been a clear clue as to what the group we will be examining next is, here is the official definition

Definition. *The symmetric group for $n \geq 3$ is the set of permutations $S_n = \{\sigma : [n] \to [n]\}$ under function composition. The identity is simply the permutation that maps every element to it's original spot, essentially doing nothing. The identity is denoted id.*

There are two ways to express the elements of S_n. With two row rotation, we list all n integers in the first row and the results of the permutation on the second row, for $\sigma \in S_n$

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ \sigma(1) & \sigma(2) & \sigma(3) & \cdots & \sigma(n) \end{pmatrix}$$

Definition. *A cycle is a string of integers that represent the items being permuted. As the name suggests, every integer that appears in the cycle is cyclically permuted while every other integer is held in place. For instance the cycle $(1\,2\,3)$ would map 1 to 2, 2 to 3 and 3 to 1, while every other element stays in place.*

Every elements in the symmetric group S_n can be expressed as a product of disjoint cycles, these cycles can be found using cycle decomposition.

Example.

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 1 & 2 \end{pmatrix} = (1\,3\,2\,4)$$

In the general case, cycle decomposition is done as follows

1. Start a cycle at some n (usually 1 for convenience)
2. Find $\sigma(n)$ and record it next
3. Continue until the cycle ends and we reach n again
4. Choose another n not previously seen and repeat until every integer is recorded

Example.

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 7 & 8 & 3 & 1 & 2 & 9 & 5 & 10 & 6 & 4 \end{pmatrix} = (1\,7\,5\,2\,8\,10\,4)(6\,9)$$

By convention, single element cycles are omitted when performing cycle decomposition. It is possible to express every element in S_n as the product of disjoint cycles up to permutation.

To take the product of two permutations, first express both permutations in terms of cycles and conjugate them together. To determine where an element gets permuted to, we start by first permuting it using the right most cycle. If the resulting integer appears in any of the cycles to the left, we follow it to said cycle and permute again. The process is repeated until the final integer is either at the left-most cycle or no longer appears in any more cycles.

Example. Let $\sigma_1 = (1\,2)(3\,4\,5)$ and $\sigma_2 = (2\,4\,1\,5)$, then

$$\sigma_1\sigma_2 = (1\,2)(3\,4\,5)(2\,4\,1\,5) = (1\,3\,4\,2\,5)$$

Example. $(1\ \ 2\ \ 3)(1\ \ 2)(3\ \ 4) = (1\ \ 3\ \ 4)$

Example. $(1\ \ 2)(1\ \ 3) = (1\ \ 3\ \ 2) \qquad (1\ \ 3)(1\ \ 2) = (1\ \ 2\ \ 3)$

The last example shows that multiplication is not commutative and so the symmetric group is not abelian. It's easy to see from how multiplication is carried out that only disjoint cycles commute.

To take inverses in the symmetric group, we must first find how to invert a cycle. Intuitively, if we were to cyclically permute some number of integers then the "undo" button would just be to shuffle the integers backwards along the same cycle. Mathematically this means that

$$(a_1\,a_2\,\ldots\,a_m)^{-1} = (a_m\,\ldots\,a_2\,a_1)$$

We can now use the fact that every permutation can be cyclically decomposed and the "shoes and socks" theorem to invert any permutation. Since the cycles in a cyclic decomposition are disjoint, we can effectively ignore the product reversing nature of inversion and so

$$\sigma^{-1} = \tau_k^{-1}\cdots\tau_1^{-1} = \tau_1^{-1}\cdots\tau_k^{-1}$$

Intuitively, if we had a cycle of n integers and we shuffled all of them around the cycle n times, then we would get to our original result. The implication of this of course is that the order of a n-cycle is n. We note that shuffling the integers around by any multiple of n would also yield the same result. If we had a permutation made up of disjoint cycles, then to reach the original starting permutation we must apply each cycle some number of times equal to a multiple of its order. Thus if a permutation had order m, then m must be divisible by the lengths of every constituent cycle. We will express these observations in the following proposition.

Proposition 2.2. We can determine the order of any element in the symmetric group using two principles

1. The order of a cycle is its length

$$|(a_1\, a_2\, \ldots\, a_m)| = m$$

2. The order of a permutation is the LCM of its constituent cycles

$$\sigma = \tau_1 \cdots \tau_k \to |\sigma| = lcm(|\tau_1|, \ldots, |\tau_j|)$$

Example. —(1 2)(3 4 5)— = lcm(2, 3) = 6

Since there are $n!$ ways to permute n elements we note that $|S_n| = n!$

Definition. *The dihedral group D_{2n} is the set of symmetries of a regular n-gon. Every element in the dihedral group is expressed in terms of rotations and reflections.*

There are two basic elements in the dihedral group: r, which is a counterclockwise rotation of the vertices and s, which is reflection about the center axis. There are a total of n rotations and n reflections and so

$$D_{2n} = \left\{1, r, r^2, \ldots, r^{n-1}, s, sr, \ldots, sr^{n-1}\right\} \to |D_{2n}| = 2n$$

Proposition 2.3 (Properties of the Dihedral Group)**.**
There are a few key properties of D_{2n}

1. $|r| = n \to r^{-k} = r^{n-k}$
2. $|s| = 2 \to s^{-1} = s$
3. $sr^k = r^{-k}s$ and so D_{2n} is non-abelian
4. $|sr^k| = 2$

Defining the identity to be 1, the n-gon without any rotations nor reflections, along with the above properties show that the dihedral group is indeed a group.

We note that every element in D_{2n} can be expressed in terms of some amount of r's and s's. Thus we can say that D_{2n} is generated by r and s.

Definition. *A group G is generated by a subset S is every element of G can be expressed as a product of elements in S. We can expressed a group using its generators through a group presentation*

$$G = \langle S \mid \textit{relations between generators} \rangle$$

The group presentation of the dihedral group is thus

$$D_{2n} = \langle r, s \mid r^n = 1\, s^2 = 1\, rs = sr^{-1} \rangle$$

The next few examples are less commonly used than the previous two but still noteworthy enough to discuss. The Quaternion Group Q_8 is a group of order 8 defined by the following products

Definition. $Q_8 = \{\pm 1, \pm i, \pm j, \pm k\}$

$$\begin{gathered}(-1)a = -a = a(-1) \quad \forall a \in Q_8 \\ (-1)^2 = 1 \\ i^2 = j^2 = k^2 = -1 \\ ij = k \quad jk = i \quad ki = j\end{gathered}$$

We can also express the quaternion group in terms of its generators

$$Q_8 = \langle -1, i, j, k \mid (-1)^2 = 1,\, i^2 = j^2 = k^2 = ijk = -1 \rangle$$

The Klein 4-group is a 4 elements group in which every element has order 2. It is smallest non-cyclic group up to isomorphism and is abelian.

Definition. $K_4 = \{1, a, b, c\}$

$$\begin{gathered}a^2 = b^2 = c^2 = 1 \\ ab = ba = c \quad bc = cb = a \quad ac = ca = b\end{gathered}$$

It can also be presented in terms of any two non-identity elements

$$K_4 = \langle a, b \mid a^2 = b^2 = (ab)^2 = 1 \rangle$$

Similar to the dihedral group is the group of rigid motions, also known as the group of isometries on $\mathbb{R}^n$. An isometry is any transformation preserving distance, such as translations and rotationss. This group is commonly denoted $E(n)$ and is generated by translations T_a and linear isometries Φ.

Definition. *A linear isometry is a functions f satisfies $|f(a)| = |a|$ and is linear, that is*

$$f(ab + c) = f(a)f(b) + f(c)$$

Definition. *A homomorphism is a function between two groups $\phi : G \to H$ satisfying $\phi(xy) = \phi(x)\phi(y)$. A group isomorphism is a bijective group homomorphism. Two groups are isomorphic if an isomorphism exists between them.*

Example. The identity map from any group to itself is a group isomorphism.

Example. The map $\phi : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z} \quad \phi(x) = \overline{x}$ is a homormorphism. It is surjective but no injective and thus not an isomorphism.

Example. The determinant $det : GL_n(\mathbb{Q}) \to \mathbb{Q}^\times$ is a group homomorphism.

Proposition 2.4 (Properties of Group Homomorphisms)**.**
Let $\phi : G \to H$ be a group homomorphism and let it be an isomorphism in (2)-(4)

1. $\phi(1_G) = 1_H \qquad \phi(x^{-1}) = (\phi(x))^{-1}$
2. ϕ^{-1} is a group homomorphism
3. $|x| = |\phi(x)|$
4. G is abelian if and only if H is abelian

Thus we can imagine an isomorphism as a map preserving group structure.

Example. There are three groups of order 8: $\mathbb{Z}/8\mathbb{Z}$, D_8, Q_8 and these are not isomorphic. $\mathbb{Z}/8\mathbb{Z}$ is abelian while the other two are not and thus cannot be isomorphic to them. D_8 has five elements of order 2 while Q_8 only has one element of order 2 and so they also cannot be isomorphic. Thus these groups have unique structures.

Similar to how we may discuss domain and range when considering normal functions in $\mathbb{R}^n$, we can define the kernel and image of a homomorphism.

Definition. *Let $\phi : G \to H$ be a group homomorphism, the kernel is defined*

$$\ker \phi = \{g \in G \mid \phi(g) = 1_H\} \subseteq G$$

The image is analogous to the range of a homomorphism and is defined

$$im\ \phi = \{h \in H \mid \exists g \in G \ s.t.\ \phi(g) = h\} = \{\phi(g) \mid g \in G\} \subseteq H$$

2.2 Subgroups

Definition. *Let G be a group and H be a nonempty subset of G. H is a subgroup of G if it is closed under multiplication and inversion, forming a group with the same group operation as G. A subgroup is denoted $H \leq G$.*

Example. The following are all subgroups

1. $\{1_G\}, G \leq G$
2. $\mathbb{Z} \leq \mathbb{Q} \leq \mathbb{R} \leq \mathbb{C}$
3. $\mathbb{Q}^\times \leq \mathbb{R}^\times \leq \mathbb{C}\times$
4. $S_m \leq S_n$ if $m \leq n$

Proposition 2.5 (The Subgroup Criterion)**.**
Let G be a group and $H \subseteq G$ be a subset. $H \leq G$ iff

1. $H \neq \varnothing$
2. $x, y \in H \to xy^{-1} \in H$

Proof. Suppose H is a nonempty subset of G satisfying the subgroup criterion. Let $x = 1$ and $y \in H$ be any elements. The criterion thus reads $1 \cdot y^{-1} = y^{-1} \in H$, showing that H is closed under inversion. Now select y^{-1} to replace y. The criterion now reads $x(y^{-1})^{-1} = xy \in H$ and so H is also closed under multiplication. Thus H is a subgroup of G.

Now suppose $H \leq G$. For any two elements $x, y \in H$ we have $xy^{-1} \in H$ because H is be definition closed under multiplication and inversion. Thus the subgroup criterion holds. □

Proposition 2.6. Let H_1, H_2 be subgroups of G.

1. $H_1 + H_2 = \{x + y \mid x \in H_1,\, y \in H_2\} \leq G$ where G is abelian under the operation $+$
2. $H_1 \cap H_2 \leq G$

The second point of the proposition can be extended further to finite intersections.

Proposition 2.7. Let $\{H_i\}_{i \in I}$ be an indexed set of subgroups $H_i \leq G$. Then

$$\bigcap_{i \in I} H_i \leq G$$

Definition. *A cyclic group is a group generated by a single element, for instance*

$$G = \langle x \rangle = \left\{\ldots, x^-2, x^{-1}, 1, x, x^2, \ldots\right\}$$

A cyclic subgroup is the subgroup generated by a single element of the group

$$\langle x \rangle = \left\{\ldots, x^-2, x^{-1}, 1, x, x^2, \ldots\right\} \leq G$$

Example. The integers form a cyclic group generated by 1

$$\mathbb{Z} = \langle 1 \rangle = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$$

The integers mod n is also a cyclic group generated by $\overline{1}$

$$\mathbb{Z}/n\mathbb{Z} = \langle \overline{1} \rangle = \left\{\overline{1}, \overline{1+1} = \overline{2}, \overline{1+1+1} = \overline{3}, \ldots, \overline{n-1}\right\}$$

The above two are the only two cyclic groups up the isomorphism, that is every other cyclic group is isomorphic to one of the above two groups. Every infinite cyclic group is isomorphic to $\mathbb{Z}$ and every finite cyclic group is isomorphic to $\mathbb{Z}/n\mathbb{Z}$.

Proposition 2.8. Let G be a group and $x \in G$ such that $|x| = n$

1. $x^k = 1 \to n|k$
2. $\langle x \rangle = \{1, x, \ldots, x^{n-1}\} \cong \mathbb{Z}/n\mathbb{Z}$
3. $|x^a| = \dfrac{n}{gcd(a,n)}$

Proof. We will prove these one at a time

1. Let $d = gcd(n,k)$ and use the Euclidean Algorithm to decompose as $d = an + bk$. Thus
$$x^d = x^{an+bk} = (x^n)^a(x^k)^b = 1^a 1^b = 1$$
By definition, d is the greatest integer dividing both n and k and so $0 < d \leq n$. But n is the smallest integer such that $x^n = 1$, therefore it must be the case that $d = n$. Thus $n|k$, as expected.

2. $\langle x \rangle = \left\{x^k \mid k \in \mathbb{Z}\right\}$. We can use the division algorithm to get $k = nq + r$. Thus using point (1)
$$x^k = x^{nq+r} = x^n q x^r = x^r \in \left\{1, x, \ldots, x^{n-1}\right\}$$
Thus there are only n distinct elements in $\langle x \rangle$ and so $|\langle x \rangle| = n$. Therefore we can define the isomorphism $\phi : \langle x \rangle \to \mathbb{Z}/n\mathbb{Z}\, \phi(x^k) = \overline{k}$

3. Let $d = gcd(a.n)$ and rewrite $a = pd, n = qd$. Since we defined d to be the GCD, this forces $gcd(p,q) = 1$, otherwise d would not be the greatest factor. Suppose $|x^a| = m$.
$$(x^a)^q = x^{qa} = x^{qpd} = x^{np} = (x^n)^p = 1^p = 1 \to |x^a| = q \to m|q$$
The last point follows by using the first part of this proposition.
$$(x^a)^m = x^{am} = 1 \to n|am \to qd|pdm \to q|pm$$
But p and q are coprime, so this can only hold if $p|m$. However we established previously that $m|q$, thus it must be that $m = q$.
$$m = q = \frac{n}{d} = \frac{n}{gcd(a,n)}$$

□

Given a group with known order, it is very easy to determine if said group is cyclic using the following proposition.

Proposition 2.9. Let G be a group of order n, then G is cyclic if and only if there exists an element of G with order n.

Proof. Suppose G is cyclic and generated by x, then $|x| \leq |G|$ because the order of a group cannot be less than the order of any of its elements. But for x to generate every element of G, it must have order at least n, $|x| \geq n$. Thus $|x| = n$.

Conversely suppose $|G| = |x| = n$, then $\langle x \rangle = \{1, x, \ldots, x^{n-1}\} \leq G$. But this means that $|\langle x \rangle| = |G| = n$, so $G = \langle x \rangle$ and so G is cyclic. □

We can extend the notion of a cyclic subgroup to subgroups generated by more than one element, which contain all possible products of the generating set.

Definition. *If G is a group and A a nonempty subset of G, then the subgroup of G generated by A is the group*

$$\langle A \rangle = \left\{ \prod_i a_i \,\middle|\, a_i \text{ or } a_i^{-1} \in A \right\} \leq G$$

This can also be defined as the intersection of all subgroups of G containing A

$$\langle A \rangle = \bigcap_{A \subseteq H_i \leq G} H_i$$

There a few very important families of subgroups that we will examine in further detail. For the following discussion let G be a group and $A \subseteq G$ any nonempty subset.

Definition. *The centralizer of A in G is the set*

$$C_G(A) = \left\{ g \in G \mid gag^{-1} = a \, \forall a \in A \right\}$$

The center of a group G is the set

$$Z(G) = C_G(G) = \{ g \in G | ga = ag \, \forall a \in G \}$$

Note that the condition $gag^{-1} = a$ implies $ga = ag$ and so the centralizer is the set of elements of G that commute with every element of A. Similarly the center of the group would be the set of elements that commute with every other elements in that group.

Definition. *The normalizer of A in G is the set*

$$N_G(A) = \left\{ g \in G \mid gAg^{-1} = A \right\}$$

where $gAg^{-1} = \{gag^{-1} \mid a \in A\}$

The normalizer condition can be written as $gA = Ag$, which is a weaker condition than the centralizer condition. Essentially these are the elements that "roughly commute" as acting on the entire set from the left and right just needs to yields the same results.

Proposition 2.10. Consider the sets defined above

1. $C_G(A) \leq N_G(A) \leq G$
2. If $H \leq G$, then $H \leq N_G(H)$. Specifically $N_G(G) = G$
3. Suppose $H \leq G$ and G is abelian, then $H \leq C_G(H)$

Proof. We will prove these one at a time

1. First note that $C_G(A) \subseteq N_G(A) \subseteq G$. Let $x, y \in C_G(A)$ and so $xax^{-1} = a$, $yay^{-1} = a$. Acting on the left with x^{-1} and on the right with x yields $a = x^{-1}ax \to x^{-1} \in C_G(A)$. Now consider the product

$$\begin{aligned}(xy)a(xy)^{-1} &= (xy)a(y^{-1}x^{-1}) \\ &= x(yay^{-1})x^{-1} \\ &= xax^{-1} = a\end{aligned}$$

   Thus $xy \in C_G(A)$ and so the centralizer is a subgroup of G. An analogous proof applies to $N_G(A)$. Thus $C_G(A), N_G(A) \leq G \to C_G(A) \leq N_G(A) \leq G$.

2. Consider some element $h \in H$, since H is closed under inversion $h^{-1} \in H$ and so $hHh^{-1} = H \to h \in N_G(H)$. Thus it follows that $G \leq N_G(G) \leq G \to N_G(G) = G$.

3. Consider some element $h \in H$. Since G is abelian and thus so is H,

$$\forall a \in H, hah^{-1} = h(h^{-1}a) = (hh^{-1})a = a$$

   Since H is abelian, we expect that every element of H lies in the centralizer which is indeed the case we see here. Thus $H \leq C_G(H)$.

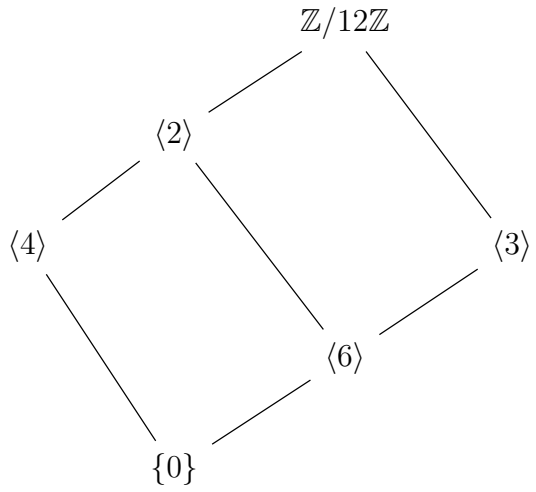
Note that in general it is not the case that $H \leq C_G(H)$. □

Since the center of G is just the centralizer of G in G, the center is also a subgroup $Z(G) \leq G$.

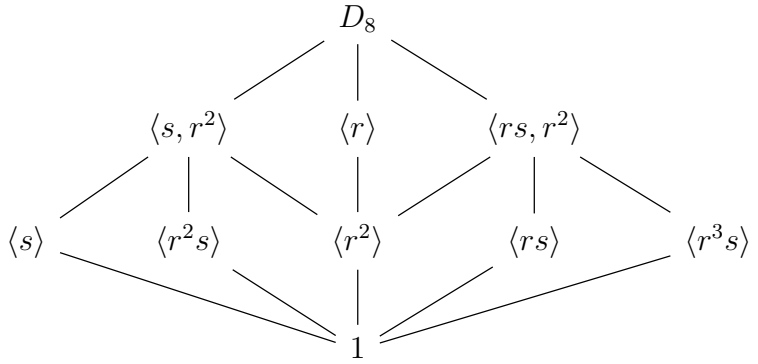
We can visualize the subgroups of a group using a diagram called a lattice. The construction of lattice diagrams will play an important role in understanding Galois theory later.

To construct a lattice diagram we start with the trivial subgroups $1, G$ on the bottom at top respectively. The other subgroups are plotted in between in roughly increasing order of size. Two subgroups A, B are connect by a line if $A \leq B$. Note that this allows us to find the smallest subgroup containing two smaller groups, namely the first common group we hit if we traced paths upwards. Similarly, we can find the largest subgroup contained within two other groups. Two identical lattices implies that two groups are isomorphic.

Example. The lattice diagram for $\mathbb{Z}/12\mathbb{Z}$

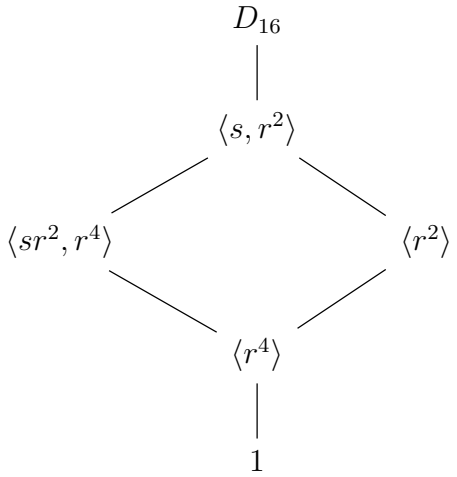


Example. The lattice of the dihedral group D_8 is



For larger groups the lattice diagram can become a bit unwieldy with crossed lines and whatnot. Usually we are only interested in a select few subgroups. In these cases we can plot just those groups in what is known as a sublattice.

Example. Suppose we only cared about the subgroups $\langle sr^2, r^4 \rangle, \langle r^2 \rangle$ of D_{16}, the sublattice is



2.3 Quotient Groups

Definition. *For a group G, subgroup $H \leq G$, and $a \in G$. The left coset is the set*

$$aH = \{ah \mid h \in H\}$$

Similarly, the right coset is defined as

$$Ha = \{ha \mid h \in H\}$$

In an abelian group, the left and right cosets are equal.

Example. Let $G = \mathbb{Z}$ and consider the subgroup $n\mathbb{Z}$ for some integer n. Then for any integer k, the left coset is

$$k + n\mathbb{Z} = \{k + nq \mid q \in \mathbb{Z}\}$$

Example. Let $G = D_8$ and consider the subgroup generated by r, $H = \langle r \rangle$. We have the following cosets

$$\begin{aligned} 1H = H &= \left\{1, r, r^2, r^3\right\} \\ rH &= \left\{r, r^2, r^3, 1\right\} = H \\ sH &= \left\{s, sr, sr^2, sr^3\right\} \\ srH &= \left\{sr, sr^2, sr^3, s\right\} = sH \end{aligned}$$

We note that $H = rH = r^2H = r^3H$ and $sH = srH = sr^2H = sr^3H$

There are a few observations we can make from the above examples

1. Two cosets are either equal or disjoint
2. The union of all cosets is the full group G
3. All cosets of a subgroup have the same size (cardinality)

From these observations we can conclude that the cosets partition the group G, furthermore the partitions are all equal with size $|H|$. We would like to use these observations to define a way to divide groups, thus we will extend the modulo operator to groups.

Proposition 2.11. Let $H \leq G$ and define the following relation

$$a \equiv b \mod H \to b = ah \text{ for some } h \in H$$

mod h defines an equivalence relation and the equivalence classes are the left cosets aH

We use denote the equivalence classes of an element $a \in G$ as

$$[a] = \{b \in G \mid a \equiv b \mod H\} = \{b \in G \mid b = ah \text{ for some } h \in H\} = aH$$

Proposition 2.12. Let S be a set, $\sim$ be an equivalence relation and $[a]$ be the equivalence classes, then the following are equivalent

1. $a \sim b$
2. $b \in [a]$
3. $[a] = [b]$

Proof. 2 follows from 1 by definition.
Let $c \in [a] \to a \in c$, $b \in [a] \to a \sim b \to b \sim a$ by symmetry. Thus using transitivity gives us $b \sim c \to c \in [b]$. Therefore all elements of $[a]$ are in $[b]$ and so $[a] = [b]$ showing that 3 follows from 2
By reflexivity $b \sim b \to b \in [b] = [a] \to b \in [a] \to a \sim b$ and so 1 follows from 3. □

Applying this proposition to the modulo relation we defined early gives the following properties of cosets

Proposition 2.13. Let $H \leq G$, then the following are equal

1. $a = bh$ for some $h \in H$
2. $b \in aH$
3. $aH = bH$

Definition. *The index of a subgroup H in G is the number of left cosets of H, denoted $|G : H|$*

There is a powerful theorem relating the index of a subgroup to the orders of both the group and subgroup.

Theorem 2.14 (Lagrange)**.** Let G be a finite group and $H \leq G$. Then

$$|H| \mid |G| \text{ and } |G : H| = \frac{|G|}{|H|}$$

Proof. Let $a_1H, \ldots, a_nH$ be all the distinct cosets of some group G. All the cosets are pairwise disjoint and combine to form the total set G, thus we know that

$$G = \bigcup_{i=1}^{n} a_iH \qquad |G| = \sum_{i=1}^{n} |a_iH|$$

We also know that all the cosets have the same cardinality and so

$$|G| = n|H| \to |H| \mid |G| \to n = |G : H| = \frac{|G|}{|H|}$$

□

Corollary 2.15. If G is a finite group and $x \in G$, then $|x| \mid |G|$

Proof. $|x| = |\langle x \rangle|$. Since x generates a cyclic subgroup of G, we must have that $|x| \mid |G|$ $\square$

Corollary 2.16. Suppose G is a group of prime order $|G| = p$. Let $x \in G \setminus \{1\}$, then $G = \langle x \rangle$. In particular, $G \cong \mathbb{Z}/p\mathbb{Z}$ and G has only two subgroups $\{1\}$ and G

Proof. By Langrange's Theorem, $|x|$ must divide $|G| = p$ and since p is prime, it must be the case that $|x| = p$ since $x \neq 1$. Since both x and the group G have the same order, we can conclude that G is cyclically generated by x, that is

$$G = \langle x \rangle \cong \mathbb{Z}/p\mathbb{Z}$$

To show the second part of the corollary, suppose we have a subgroup $H \leq G$ that is not just the identity element $H \neq \{1\}$. Thus there is some non-identity element $x \in H$, but this means that $H = \langle x \rangle = G$, thus the only two possible subgroups are G and $\{1\}$. $\square$

Cosets are important in group theory because they allow us to "divide" one group by another.

Definition. *Let G be a group and $H \leq G$, then we can define the quotient $(G \mod H)$ as*

$$G/H = \{aH \mid a \in G\}$$

Thus the quotient of one group by a subgroup is just the set of all its cosets. We can see this in action and confirm that it has the desired functionality be examining the integers.

Example. Consider the subgroup $n\mathbb{Z} \leq \mathbb{Z}$ and form the quotient

$$z/n\mathbb{Z} = \{a + n\mathbb{Z} \mid a \in \mathbb{Z}\} = \{a + kn \mid a, k \in \mathbb{Z}\} = \{\overline{a} \mid a \in \mathbb{Z}\}$$

A key note from the above example is that the group operation on $\mathbb{Z}$ induces another group operation on the resulting quotient. In this case addition between residue classes is well defined, that is

$$\overline{a} + \overline{b} = \overline{a + b} \in \mathbb{Z}/n\mathbb{Z}$$

In group theoretic terms, if we have two cosets aH and bH, we would desire $aHbH = (ab)H$. In other words, we require that left and right cosets be the same for the quotient to be a group under the induced group operation. This leads us into the next definition.

Definition. *Let $H \leq G$. H is a normal subgroup if $aH = Ha$, $\forall a \in G$ denoted $H \trianglelefteq G$*

Example. Consider the dihedral group $G = D_8$ and consider the subgroup $K = \langle s \rangle = \{1, s\}$

$$rK = \{r, rs\} = \left\{r, sr^3\right\} \neq Kr = \{r, sr\}$$

Therefore K is not a normal subgroup of D_8. Now consider the cyclic group generated by r, $H = \langle r \rangle$, we know that $r^n H = Hr^n = H$, thus we just need to show that $sH = Hs$

$$sH = \left\{s, sr, sr^2, sr^3\right\} = Hs = \left\{s, rs, r^2s, r^3s\right\} = \left\{s, sr^3, sr^2, sr\right\}$$

Thus H is indeed a normal subgroup of D_8, which allows us to define the quotient G/H

There are multiple ways to characterize normal subgroups. In fact there are a whole slew of conditions, each of which would suffice to show that a subgroup is normal.

Proposition 2.17 (Internal Characterization of Normal Subgroups)**.**
Let G be a group and $H \leq G$, the following are equivalent

1. $H \trianglelefteq G$
2. $aH = Ha$, $\forall a \in G$
3. $N_G(H) = G$
4. $aHa^{-1} = H$, $\forall a \in G$
5. $aHa^{-1} \subseteq H$, $\forall a \in G$
6. $aha^{-1} \in H$, $\forall a \in G, \forall h \in H$

Proof. $2 \leftrightarrow 1$ by definition.
$2 \leftrightarrow 4$ by using some algebraic manipulation $3 \leftrightarrow 4$ by using the definition of the normalizer $6 \leftrightarrow$ by using the definition of subsets.
$4 \rightarrow 5$ is trivial as if two sets are equal than each one is a subset of the other.
$5 \rightarrow 4$ is slightly more difficult. Let $a \in G$ and since a group is closed under inversion, point 5 states that $aHa^{-1}, a^{-1}Ha \subseteq H$, thus

$$H = a(a^{-1}Ha)a^{-1} \subseteq aHa^{-1} \subseteq H \rightarrow H \subseteq aHa^{-1}$$

Thus we must have that $aHa^{-1} = H$ □

Example. Consider the subgroup $SL_n(\mathbb{R}) \leq GL_n(\mathbb{R})$. Let $A \in GL_n(\mathbb{R})$ and $B \in SL_n(\mathbb{R})$, then

$$det(ABA^{-1}) = det(A)det(B)det(A^{-1}) = det(B) = 1$$

Thus $ABA^{-1} \in SL_n(\mathbb{R})$ and so $SL_n(\mathbb{R}) \trianglelefteq GL_n(\mathbb{R})$

Proposition 2.18. Here we will highlight some useful properties of normal subgroups

1. $\{1\}, G \trianglelefteq G$
2. $H \trianglelefteq G$, $H \leq K \leq G \rightarrow H \trianglelefteq K$
3. $H \trianglelefteq H_G(H)$, $H \trianglelefteq K \rightarrow K \leq N_G(H)$
4. $Z(G) \trianglelefteq G$. If $H \trianglelefteq Z(G)$, then $H \trianglelefteq G$
5. $|G : H| = 2 \rightarrow H \trianglelefteq G$

Proof. We'll prove these one at a time

1. Trivial

2. For any $k \in K \subseteq G$, $kH = Hk$ since $H \trianglelefteq G$, thus $H \trianglelefteq K$

3. Let $g \in N_G(H)$, which means $gHg^{-1} = H \to H \trianglelefteq N_G(H)$. If $H \trianglelefteq K$, then $\forall k \in K$, $kHk^{-1} = H$, thus $K \leq N_G(H)$

4. Let $H \leq Z(G)$

$$\begin{aligned} aH &= \{ah \mid \forall h \in H\} \\ &= \{ha \mid \forall h \in H\} \\ &= Ha, \forall a \in G \end{aligned}$$

   Thus $H \trianglelefteq G$

5. If the number of cosets is 2, that means the only left and right cosets are $H, G \setminus H$. Thus for any element $a \in G$, there are two cases

$$\begin{cases} a \in H & \to \quad aH = H = Ha \\ a \notin H & \to \quad aH = G \setminus H = Ha \end{cases}$$

   Thus $H \trianglelefteq G$

□

Note that in general $\trianglelefteq$ is not transitive, unlike subgroups. Thus

$$A \trianglelefteq B \trianglelefteq C \not\Rightarrow A \trianglelefteq C$$

Example. Every subgroup of Q_8 is normal

Order	Subgroups	Proof
1	$\{1\}$	Point 1
2	$\langle -1 \rangle$	± 1 commute with everything
4	$\langle i \rangle, \langle j \rangle, \langle k \rangle$	All are index 2
8	Q_8	Point 1

Now that we've established that the quotient group is indeed a group and is well defined, we might want to know if there is a "natural" map from a group to its quotient.

Definition. *Left G be a group and $H \trianglelefteq G$. The structure homomorphism for quotients is defined as*

$$\pi : G \to G/H \qquad \pi(a) = aH$$

Proposition 2.19. There are two notable properties of the structure homomorphism

1. π is surjective

2. $\ker \pi = H$

Before continuing with our discussion of quotient groups there are a few properties of homomorphisms that will be useful that we will note here.

Proposition 2.20. Let $\phi : G \to H$ be a group homomorphism.

1. $\ker \phi \trianglelefteq G$
2. $\text{im } \phi \leq H$
3. ϕ is injective if and only if $\ker \phi = \{1_G\}$

Proof. We will prove these separately

1. The kernel cannot be empty because the identity is always preserved. Let $a, b \in ker(\phi)$, thus $\phi(a) = \phi(b) = 1_H$ and so
$$\phi(ab^{-1}) = \phi(a)\phi(b)^{-1} = 1_H 1_H^{-1} = 1_H$$
and so the kernel is a subgroup. Now consider some other element $g \in G$,
$$\phi(gag^{-1}) = \phi(g)\phi(a)\phi(g)^{-1} = \phi(g)\phi(g)^{-1} = \phi(gg^{-1}) = \phi(1_G) = 1_H$$
Thus $gag^{-1} \in$
$ker(\phi)$ and so it is a normal subgroup
$ker(\phi) \trianglelefteq G$

2. Since G is non-empty, the image must also be non-empty. Let $x, y \in$
$im(\phi)$ and suppose $\phi(a) = x, \phi(b) = y$. Then
$$\phi(ab^{-1}) = \phi(a) \in \phi(b) = xy^{-1} \in im(\phi)$$
Thus the image is a subgroup im $(\phi) \leq H$

3. Suppose ϕ is injective and let $a \in$
$ker(\phi)$. The homomorphism must preserve the identity element and since ϕ is injective, it must be that $a = 1_G$, thus
$ker(\phi) = \{1_G\}$.
Conversely suppose
$ker(\phi) = \{1_G\}$ and $\phi(a) = \phi(b)$, then
$$\phi(a)\phi(b)^{-1} = \phi(ab^{-1}) = 1_H \to ab^{-1} = 1_h \to a = b$$
Therefore ϕ is injective.

□

The first point gives us a way to externally characterize normal subgroups.

Proposition 2.21. Let $H \leq G$, then $H \trianglelefteq G$ if and only if $H =$
$ker(\phi)$ for some homomorphism $\phi : G \to G'$

Proof. If $H \trianglelefteq G$, then H is the kernel of the structure homomorphism $\pi : G \to G/H$. Conversely if $H =$
$ker(\phi)$, then $H \trianglelefteq G$ as proved above. □

Example. $SL_n(\mathbb{R}) \trianglelefteq GL_n(\mathbb{R})$ because $SL_n(\mathbb{R})$ is the kernel of the determinant function.

2.4 The Isomorphism Theorems

The four isomorphism theorems are a series of powerful theorems relating homormorphism and quotient groups.

Theorem 2.22 (First Isomorphism Theorem)**.**
Let $\phi : G \to H$ be a group homomorphism, then

$$G/ker(\phi) \cong im(\phi)$$

Proof. Let $K =$
$ker(\phi)$ as by the proposition. Define the homomorphism

$$\overline{\phi} = G/K \to K \qquad \overline{\phi}(ak) = \phi(a)$$

Consider some element $ak \in$
$ker(\overline{\phi})$, so that $\overline{\phi}(ak) = \phi(a) = 1_H$, therefore

$$a \in K = \ker(\phi) \to aK \subseteq K \to aK = K$$

Thus
$ker(\overline{\phi}) = \{1K\}$ which is the identity in the quotient group, thus $\overline{\phi}$ is injective. We can easily make $\overline{\phi}$ surjective by restricting the codomain to the image of ϕ is other words, the homomorphism $\overline{\phi} : G/K \to$
$im(\phi)$ is bijective and thus is an isomorphism. Therefore

$$G/ker(\phi) \cong im(\phi)$$

as required □

Example. $GL_n(\mathbb{R})/SL_n(\mathbb{R}) \cong \mathbb{R}^\times$ using the determinant function.

Example. $\mathbb{R}^\times/\mathbb{R}_+ \cong \mathbb{Z}/2\mathbb{Z} \cong \{\pm 1\}$ using the map

$$\phi(x) = \frac{x}{|x|} = \begin{cases} -1 & x > 0 \\ 1 & x < 0 \end{cases}$$

Theorem 2.23 (Second (Diamond) Isomorphism Theorem)**.** Let $A, B \leq G$ and $A \leq N_G(B)$, then

$$AB = \{ab \mid a \in A,\, b \in B\} \leq G$$

Furthermore $B \trianglelefteq AB$, $A \cap B \trianglelefteq A$ and $AB/B \cong A/A \cap B$

Before we prove the second isomorphism theorem, we will first prove a helpful proposition.

Proposition 2.24. If $A, B \leq G$ and $A \leq N_G(B)$, then $AB \leq G$.

Proof. Because both A, B are nonempty, AB is also nonempty. Consider some element $xy \in AB$ where $x = a_1b_1$, $y = a_2b_2$, then $xy^{-1} = a_1b_1b_2^{-1}a_2^{-1}$. Because $A \leq N_G(B)$, $a_2Ba_2^{-1} = B$. Define $b_3 = a_2b_1b_2^{-1}a_2^{-1} \in B$, thus

$$xy^{-1} = a_1b_1b_2^{-1}a_2^{-1} = a_1a_2^{-1}b_3 \in AB$$

and so $AB \leq G$ □

The proposition proves the first part of the second isomorphism theorem, we can now proceed to prove the rest of the theorem.

Proof. We'll start by proving $B \trianglelefteq AB$. Consider some $x \in AB, y \in B$ and suppose $x = ab$ where $a \in A, b \in B$. Then $xyx^{-1} = abyb^{-1}a^{-1}$, let $z = byb^{-1} \in B$. Then $xyx^{-1} = aza^{-1} \in B$ because $A \leq N_G(B)$ and so $B \trianglelefteq AB$.

Now we will show that $a \cap B \trianglelefteq A$. Consider a map $\phi : A \to AB/B$ defined by $\phi(a) = aB$.

$$\phi(xy) = xyB = xByB\phi(x)\phi(y)$$

and so ϕ is a homomorphism. Suppose $a \in \ker(\phi)$, then $\phi(a) = 1B \to a \in B$. But $a \in A$, thus $\ker(\phi) = A \cap B$. Since kernels of homomorphisms are normal subgroups, $A \cap B \trianglelefteq A$.

The last part of the theorem follows directly from the first isomorphism theorem. For every $xB \in AB/B$, the preimage is $\phi(a) = aB = xB$, where $x = ab$, $a \in A$, $b \in B$. Therefore ϕ is surjective and

$$A/\ker(\phi) \cong \text{im } (\phi) \to A/A \cap B \cong AB/B$$

□

Corollary 2.25. If $|AB| < \infty$ and $A \leq N_G(B)$, then

$$|AB| = \frac{|A||B|}{|A \cap B|}$$

Proof. Using the second isomorphism theorem

$$\left|\frac{AB}{B}\right| = \left|\frac{A}{A \cap B}\right|$$

Using Lagrange's theorem gives

$$\left|\frac{AB}{B}\right| = \frac{|AB|}{|B|} \qquad \left|\frac{A}{A \cap B}\right| = \frac{|A|}{|A \cap B|}$$

Rearranging terms gives us the desired result: $|AB| = \frac{|A||B|}{|A \cap B|}$ □

Theorem 2.26 (Third Isomorphism Theorem)**.** If G is a group with normal subgroups $H, K \trianglelefteq G$ and $H \leq K$, then

$$\frac{G/H}{K/H} \cong G/K$$

Proof. Define the structure homomorphism $\pi : G \to G/K$, $\pi(a) = aK$. Consider the subgroup $H \leq K = \ker(\pi)$ and define another homomorphism

$$\overline{\pi} : G/H \to G/K \qquad \overline{\pi}(aH) = aK$$

From the definition, we see that $\overline{\pi}$ is surjective, thus

$$\forall aK \in G/K \qquad \overline{\pi}(aH) = aK$$

Consider an element in the kernel $aH \in \ker(\overline{\pi})$

$$\begin{gathered}\overline{\pi}(aH) = aK = 1K \to a \in K \\ aH \in K/H = \{kH \mid k \in K\} \\ \therefore \ker(\overline{\pi}) = K/H\end{gathered}$$

Thus by the first isomorphism theorem

$$\frac{G/H}{\ker(\overline{\phi})} = \frac{G/H}{K/H} = G/K$$

□

Corollary 2.27. Suppose $|G| < \infty$, $H, K \trianglelefteq G$, and $H \leq K \leq G$, then

$$|G : H| = |G : K||K : H|$$

Proof. By the third isomorphism theorem

$$\left|\frac{G/H}{K/H}\right| = |G/K| = |G : K|$$

Lagrange's theorem gives

$$\left|\frac{G/H}{K/H}\right| = \frac{|G/H|}{|K/H|} = \frac{|G : H|}{|K : H|}$$

Algebraic manipulation gives us the desired result $|G : H| = |G : K||K : H|$ □

Theorem 2.28 (Fourth (Lattice) Isomorphism Theorem)**.** Suppose G is a group with normal subgroup $H \trianglelefteq G$, then there is an inclusion preserving bijection

$$\{\text{Subgroups of } G/H\} \leftrightarrow \{\text{subgroups of } G \text{ containing } H\}$$

2.5 Transpositions and the Alternating Group

We previously established that any permutation can be uniquely written as the product of disjoint cycles. We can also write any permutation as the product of permutations, but this product may not be unique nor the cycles necessarily disjoint.

Definition. *A transposition is a 2-cycle in S_n*

Proposition 2.29. The transpositions of S_n generate S_n, that is

$$S_m = \langle (i\, j) \mid 1 \leq i, j \leq n \rangle$$

Proof. We can decompose any cycle as $(a_1 \cdots a_m) = (a_1\, a_m)(a_1\, a_{m-1}) \cdots (a_1\, a_2)$. Since any permutation can be written as a product of cycles, we can use the previous formula to write any permutation as a product of transpositions. □

While there are an infinite number of ways to express a permutation as a product of transpositions, the parity (even/odd) will not change.

Definition. *An even (or odd) permutation is the product of an even (or odd) number of transpositions.*

We will show that the parity of a permutation is unique throw the following proposition.

Proposition 2.30. There exists a unique group homomorphism

$$sgn : S_n \to \{\pm 1\} \qquad sgn(\tau) = -1,\ \tau \text{ a transposition}$$

This group homomorphism distinguishes between even and odd permutations

$$sgn(\sigma) = \begin{cases} 1 & \sigma \text{ even} \\ -1 & \sigma \text{ odd} \end{cases}$$

Multiplying permutations together affects the parity in the following ways

$$(even)(even) = (odd)(odd) = even$$
$$(odd)(even) = (even)(odd) = odd$$

Intuitively this makes sense because multiplying permutations adds the lengths together.

The kernel of sgn defines an group, known as the alternation group

Definition. *The alternating group A_n for $n \geq 3$ is the set of oven permutations*

$$A_n = \{\sigma \in S_n \mid sgn(\sigma) = 1 \to \sigma \text{ is even}\}$$

Example. The alternating group for $n = 3$ is

$$A_3 = \{1, (1\,2\,3), (1\,3\,2)\} \cong \mathbb{Z}/3\mathbb{Z}$$

Proposition 2.31. The alternating group has the following useful properties

1. $A_n \trianglelefteq S_n$
2. $|S_n : A_n| = 2 \to |A_n| = \frac{n!}{2}$

Proof. These properties follow from the fact that $A_n = \ker(sgn)$ and thus from the first isomorphism theorem $S_n/A_n \cong \{\pm 1\}$ □

It may be a bit tedious trying to decompose permutations into the product of transpositions in order to find its parity, thus the following proposition may be useful when calculating alternating groups.

Proposition 2.32. A permutation is odd if and only if the number of cycles of even length in its cycle decomposition is odd.

Example. The alternating group for $n = 4$ can be calculated easily using the proposition

$$A_4 = \{1, (1\,2\,3), (1\,2\,4), (1\,3\,4), (1\,3\,2), (1\,4\,2), (1\,4\,3),\\ (2\,3\,4), (2\,4\,3), (1\,3)(2\,4), (1\,4)(2\,3), (1\,2)(3\,4)\}$$

Definition. *A group G is simple if the only normal subgroups of G are $\{1\}$ and G*

Example. The following groups are simple

1. $\mathbb{Z}/p\mathbb{Z}$ for a prime p
2. A_n is simple for $n = 3$, $n \geq 5$

3 Further Topics in Group Theory

3.1 Group Actions

Definition. *Let G be a group and A be a set. A left group action is a function $G \times A \to A$, $ga \mapsto g \cdot a$ such that*

1. *$g \cdot (h \cdot a) = (gh) \cdot a$ for $g, h \in G$, $a \in A$*
2. *$1_G \cdot a = a$*

A group action is denoted $G \curvearrowright A$

Example. The general linear group acting on the matrix group through matrix multiplication is a group action

$$GL_n(\mathbb{C}) \curvearrowright M_n(\mathbb{C}) \qquad A \cdot M \mapsto AM$$

Example. There are two ways a group can act on itself $G \curvearrowright G$

1. Left Multiplication $g \cdot a \mapsto ga$
2. Conjugation $g \cdot a \mapsto gag^{-1}$

If A is a set, we can define the set

$$S_A = \{\sigma : A \to A \mid \sigma \text{ is a bijection}\}$$

which is a group under function composition. This notation essentially extends the concept of the symmetric group to all sets. Note that $S_{\{1,2,3,\ldots,n\}} = S_n$. Using this, we can express a group action as a permutation acting on a set. Thus given a group action $G \curvearrowright A$, define

$$\sigma_g : A \to A \qquad \sigma_g(a) = ga$$

Using this defintion, we can define a homomorphism from any group G onto the symmetric group S_A

Proposition 3.1. Given a group action $G \curvearrowright A$, we can define the group homomorphism

$$\phi : G \to S_A \qquad \phi(g) = \sigma_g$$

There are a few more useful definitions to note.

Definition. *Given a group action $G \curvearrowright A$ and an element $a \in A$. The stabilizer is the set of elements that fix a*

$$G_a = \{g \in G \mid g \cdot a = a\} \leq G$$

A group action is free if the every stabilizer is trivial: $G_a = \{1\}$, $\forall a$

Example. Consider the symmetric group acting on the set $\{1, 2, 3, ..., n\}$ according to $\sigma \cdot i = \sigma(i)$. The stabilizers are

$$G_a = \{\sigma \in S_n \mid \sigma(a) = a\} \cong S_{n-1}$$

Example. Consider the two different ways a group can act on itself $G \curvearrowright G$

1. If G acts on itself by left multiplication, then

$$G_a = \{g \in G \mid ga = a\} = \{1\}$$

   In other words $G \curvearrowright G$ through left multiplication is a free group action.

2. If G acts on itself by conjugation, then

$$G_a = \left\{g \in G \mid gag^{-1} = a\right\} = C_G(a)$$

We can interpret the kernel of a group action as the elements of the group that do nothing, that is the elements that stabilize every elements of the set being acted upon.

Definition. *The kernel of a group action is the intersection of every stabilizer*

$$\ker(G \curvearrowright A) = \{g \in G \mid g \cdot a = a, \, \forall a \in A\} = \bigcap_{a \in A} G_a$$

A group action is faithful if the kernel is trivial: $\ker(G \curvearrowright A) = \{1\}$

Example. Consider the symmetric group acting on a set of integers $S_n \curvearrowright \{1, 2, 3, ..., n\}$ given by $\sigma \cdot a = \sigma(a)$. The kernel is

$$\ker(S_n \curvearrowright \{1, 2, 3, ..., n\}) = \bigcap_{i \in \{1,2,3,...,n\}} \{\sigma \in S_n \mid \sigma(i) = i\} = \{1\}$$

Thus this is a faithful group action. However as noted in a previous example, the stabilizers of this action are isomorphic to S_{n-1} and thus not trivial. Therefore this is an action that is faithful but not free.

We can use the concept of stabilizers to prove a useful theorem about the action $G \curvearrowright G$ by left multiplication.

Theorem 3.2 (Cayley)**.**
Suppose G is a finite group of order n, then $G \leq S_n$

Proof. Consider the action $G \curvearrowright G$ by left multiplication. We note that all stabilizers are trivial and so left multiplication is a faithful group action. Thus the homomorphism $\phi : G \to S_G \cong S_n$ is injective because $\ker(G \curvearrowright A) = \ker(\phi) = \{1\}$. Thus $G \cong \text{im } (\phi) \leq S_n$ □

A group action partitions the set under the group elements acting on a specific element.

Definition. *Given a group action $G \curvearrowright A$ and an element $a \in A$, the orbit of a is the set*

$$O_a = \{g \cdot a \mid g \in G\}$$

Example. Consider the action $\mathbb{R}^\times \curvearrowright \mathbb{R}^{n+1}$ given by

$$\lambda \vec{z} \mapsto (\lambda z_1, \ldots, \lambda z_n)$$

For any nonzero vector, it's ordit is

$$O_z = \left\{\lambda \vec{z} \mid \lambda \in \mathbb{R}^\times\right\}$$

which is a line in n-dimensional space. The orbit of the zero vector is the origin.

The orbits of an element O_a are the equivalence classes under the relation

$$a \sim b \to b = g \cdot a \text{ for some } g \in G$$

and thus partition the set A into disjoint subsets. If a group action doesn't partition a set, that is there is only one equivalence class containing everything, then it is said to be transitive.

Definition. *A group action is transitive if it only has one orbit*

Example. Consider the action $S_n \curvearrowright \{1, 2, 3, ..., n\}$ given by $\sigma \cdot i = \sigma(i)$. This action is transitive because for any pair $i, j \in \{1, 2, 3, ..., n\}$, we can simply select the transposition $\sigma = (i\, j) \to \sigma \cdot i = j$. Thus every pair of elements are in the same orbit and therefore every element must be in the same orbit.

Theorem 3.3 (Orbit-Stabilizer)**.**
Consider the action $G \curvearrowright A$ and some element $a \in A$, then $|O_a| = |G : G_a|$ In particular, we have a bijective map

$$G/G_a \to A \qquad gG_A \to ga$$

If G is finite, then $|G| = |O_a||G_a|$

Proof. Consider the quotient group $G/G_a = \{gG_a \mid g \in G\}$ and define the homomorphism

$$\phi : G/G_a \to O_a \qquad \phi(gG_a) = g \cdot a$$

Suppose $\phi(xG_a) = \phi(yG_a)$, then

$$x \cdot a = y \cdot a \to a = (x^{-1}y)a \to x^{-1}y \in G_a \to y \in xG_a \to yG_a = xG_a$$

Therefore ϕ is injective. To show that ϕ is also surjective, for any $g \cdot a \in O_a$, we have $\phi(gG_a) = g \cdot a$, which means ϕ is an isomorphism. Thus $|O_a| = |G/G_a|$ and Lagrange's theorem gives us the desired results. □

We've noted that there are two ways a group can act on itself. One is by regular left multiplication and the other is through conjugation. We will examine this group action in more detail.

Definition. *Two elements $a, b \in G$ are said to be conjugate if there exists another element $g \in G$ such that $gag^{-1} = b$ (and so they are in the same orbit). The orbits of $G \curvearrowright G$ by conjugation are known as the conjugacy classes of G.*
Two subsets $S, T \subseteq G$ are said to be conjugate if $gSg^{-1} = T$. In other words, the subsets belong to the same orbit.

The conjugacy classes of a group G allows us to derive a relation among the orders of each conjugacy class along with the center $Z(G)$.

Theorem 3.4 (Class Equation)**.**
Let G be a finite group and let $g_1, \ldots, g_r$ be representatives of the distinct conjugacy classes of G not contained in the center $Z(G)$. Then

$$|G| = |Z(G)| + \sum_{i=1}^{r} |G : C_G(g_i)|$$

Proof. Any element in the center $x \in Z(G)$ belongs to a conjugacy class of order 1 because $gxg^{-1} = x$ for all $g \in G$ by definition. Let $K_1, \ldots, K_r$ be the conjugacy classes not contained in the center with representative $g_i \in K_i$. Since the conjugacy classes partition the group, their orders must sum to the order of the group G, thus

$$|G| = |Z(G)| \cdot 1 + \sum_{i=1}^{r} |K_i|$$

Using the orbit stabilizer theorem and noting that $G_{g_i} = \{gag^{-1} \mid g \in G\} = C_G(g_i)$, we get

$$|O_{g_i}| = |K_i| = |G : C_G(g_i)|$$

Combining these together gives us the desired result □

Note that because the summands are indices of subgroups of the group G, they must divide $|G|$ by Lagrange's theorem. This restricts their values as well as those for $|Z(G)|$.

There is one important consequence of the class equation which allows us to determine whether a group will have a nontrivial center, that is if there is an element aside from the identity that commutes with every other element of the group.

Theorem 3.5. If P is a group of prime power order, then P has a nontrivial center. That is $Z(P) \neq \{1\}$

Proof. From the class equation, we have

$$|P| = |Z(P) + \sum_{i=1}^{r} |P : C_P(g_i)|$$

Since P is of prime power order, the right hand side must be divisible by p. Since by definition $C_G(g_i) \neq P$, we have that p divides the index $|G : C_G(g_i)|$. Thus for the equation to hold p must also divide $|Z(G)|$, implying that $|Z(G)| \neq 1$ and is therefore nontrivial. □

We can also apply group actions to prove a partial converse to Lagrange's theorem. A stronger partial converse are the Sylow theorems discussed in the next section.

Theorem 3.6 (Cauchy)**.** Let G be a finite group such that $p||G|$ where p is a prime. Then there exists a non-identity element $g \in G$ such that $g^p = 1$

Proof. Let X be the set of p-tuples of elements from G multiplying to 1

$$X = \{(g_1, \ldots, g_p) \mid g_1 \cdots g_p = 1\}$$

The group $\mathbb{Z}/p\mathbb{Z}$ acts on this set by cyclic permutation. We see that

$$(g_2 \cdots g_p)g_1 = g_1^{-1}g_1(g_2 \cdots g_p)g_1 = g_1^{-1}(g_1 \cdots g_p)g_1$$

In other words, cyclically permuting the multiplicands will change the product but only by conjugation. In particular for our set, the product of the tuple is 1 and so permutation has no effect.

We note that for each element of G^{p-1}, we can construct a unique element of X so $|X| = |G|^{p-1}$. Using the orbit-stabilizer theorem there is a bijection

$$O_x \to (\mathbb{Z}/p\mathbb{Z})/Stab(x)$$

where $Stab(x)$ denotes the stabilizer in this case. There are two possible cases

$$\begin{cases} Stab(x) = 1 \to |O_x| = 1 \\ Stab(x) = \mathbb{Z}/p\mathbb{Z} \to |O_x| = 1 \end{cases}$$

Let N_1 denote the number of orbits of size 1 and N_2 denote the number of orbits of size p. Then

$$|X| = N_1 + pN_p$$

since the orbits partition X. Note that $p||G|$ and so we can reduce $\mod p$ since $|x| = |G|^{p-1}$. Thus we get

$$0 \mod p = N_1 + 0 \mod p \to p|N_1$$

Furthermore $(1, \ldots, 1)$ is in an orbit of size 1 so $N_1 > 0$ so the number of elements of prime order is given by $N_1 - 1 \cong -1 \mod p$ which is nonzero. □

3.2 The Sylow Theorems

The Sylow theorems form partial converse to Lagrange's theorem and are useful towards classifying groups. Recall that Lagrange's theorem states that for a finite group G, the order of every subgroup of G must divide the order of G. The Sylow theorems show that for some prime p, there exists a special type of subgroup with order p^n, the highest power of p that divides $|G|$. Before continuing we will first provide some definitions.

Definition. *Let G be a group and p be some prime.*

1. *A p-group is a group of order p^n for some $n \geq 0$. Subgroups of G which are also p-groups are known as p-subgroups.*
2. *A Sylow p-subgroup is a subgroup of order p^n where n is the highest integer such that p^n divides the order of $|G|$. That is $|G| = p^n m$ for some integer m such that $p \nmid m$.*

The set of Sylow p-subgroups will be denoted by $Syl_p(G)$ and the number of Sylow p-subgroups will be denoted by $n_p(G)$. In the event that G is clear from context, it may be omitted and just n_p used instead.

We are now ready to present the Sylow theorems in their entirety.

Theorem 3.7 (Sylow)**.** Let G be a group of order $p^\alpha m$, where p is a prime not dividing m.

1. Sylow p-subgroups exist, that is

$$Syl_p(G) \neq \varnothing \to n_p(G) \neq 0$$

2. If P is a Sylow p-subgroup of G and Q is any p-subgroup, then there exists some $g \in G$ such that $Q \leq gPg^{-1}$. That is, Q is contained in some conjugate of P. In particular, any two Sylow p-subgroups of G are conjugate in G.

3. The number of Sylow p-subgroups of G is of the form $1 + kp$. Furthermore n_p is the index of the normalizer $N_G(P)$ in G for any Sylow p-subgroup P, thus n_p divides m. Mathematically that is

$$n_p \equiv 1 \mod p \qquad n_p = |G : N_G(P)| \qquad n_p | m$$

Proving the Sylow theorems is a difficult task so we will start by proving a helpful lemma.

Lemma 3.8. Let $P \in Syl_p(G)$. If Q is any p-subgroup of G, then $N_G(P) \cap Q = P \cap Q$

Proof. Let $H = N_G(P) \cap Q$. Since $P \leq N_G(P)$, it's clear that $P \cap Q \leq H$. We already know that $H \leq Q$, so to show the reverse inclusion, we just need to show that $H \leq P$.

Consider the group PH. We know that $H \leq N_G(P)$ by definition so we can use Corollary 4.15 to establish

$$|PH| = \frac{|P||H|}{|P \cap H|}$$

But every number here is a power of p and so PH must be a p-group. Furthermore $P \leq PH$ so $|PH|$ must be divisible by $|P|$. Since P is a Sylow p-group, this forces

$$|P| = |PH| \rightarrow PH = P \rightarrow H \leq P$$

Thus $H \leq P \cap Q$, which establishes the equality. □

With this lemma in hand we can now prove the Sylow theorems

Proof. (Sylow 1) This will be a proof by induction, the $|G| = 1$ case is obvious as there is nothing to prove. Assume inductively that Sylow $p-$-groups exist for all groups of order less than $|G| = p^\alpha m$.

If p divides $|Z(G)|$, then by Cauchy's theorem (5.6), there exists an element of order p which will generate a subgroup $N \leq Z(G)$ of order p. Let us define $\overline{G} = G/N$ so that $|\overline{G}| = p^{\alpha-1}m$. Since we've already assumed that every group of order less than $|G|$, $\overline{G}$ has a Sylow p-subgroup $\overline{P}$ of order $p^{\alpha-1}$. Using the Fourth Isomorphism Theorem, let P be the subgroup of G containing N such that $P/N = \overline{P}$, then

$$|P| = |P/N| \cdot |N| = p^\alpha$$

and thus P is a Sylow p-subgroup.

Now suppose p does not divide $|Z(G)|$, we have the class equation

$$|G| = |Z(G)| + \sum_{i=1}^{r} |G : C_G(g_i)|$$

where $g_1, \ldots, g_r$ are representatives of the conjugacy classes of G not contained in the center $Z(G)$. Now we note that since p divides $|G|$, it must be that there exists some i such that p does not divide $|G : C_G(g_i)|$. If p divides $|G : C_G(g_i)|$ for all i then this would also imply that p divides $|Z(G)|$, a contradiction. Let $H = C_G(g_i)$ for that such i and note

$$|H| = p^\alpha k \qquad p \nmid k$$

Since $g \notin Z(G)$, $|H| < |G|$ and so must have a Sylow p-subgroup P by induction, but

$$|P| = p^\alpha \qquad P \leq H \leq G$$

thus P is a Sylow p-subgroup, which completes the induction. □

Before we continue with the proof, lets make a few calculations. We just proved that Sylow p-subgroups exist, let $P \leq G$ be this subgroup. Furthermore let

$$\mathcal{S} = \{P_1, P_2, \ldots, P_r\} = \left\{gPg^{-1} \mid g \in G\right\}$$

be the set of all conjugates of P and let Q be any p-subgroup (not necessarily Sylow) of G. From this definition we can act on the set $\mathcal{S}$ with G (and by extension Q) by conjugation. Thus we can write

$$\mathcal{S} = O_1 \cup O_2 \cup \cdots \cup O_s$$

where O_i are the disjoint orbits and $r = |O_1| + \cdots + |O_s|$. Suppose the first s elements of $\mathcal{S}$ are representatives of the Q-orbits (we can reorder if not), so that $P_i \in O_i$. From the orbit-stabilizer theorem, we have

$$|O_i| = |Q : Stab(P_i)| = |Q : N_Q(P_i)|$$

By definition $N_Q(P_i) = N_G(P_i) \cap Q$ and using Lemma 8 we get

$$|O_i| = |Q : P_i \cap Q| \qquad 1 \leq i \leq s$$

Our choice of Q was arbitrary, so lets take $Q = P_1$ which gives

$$|O_1| = |P_1 : P_1 \cap P_1| = 1$$

For the other i's, we have $P_1 \neq P_i \to P_1 \cap P_i \leq P_1$

$$|O_i| = |P_1 : P_1 \cap P_i| > 1 \qquad 2 \leq i \leq s$$

Since P_1 is a p-group, $|P_1 : P_1 \cap P_i|$ must be a power of p and thus divisible. In other words

$$p \mid |O_i| \qquad 2 \leq i \leq s$$

This gives us the useful result

$$r = |O_1| + (|O_2| + \cdots + |O_s|) \equiv 1 \mod p$$

Now we can prove the remaining Sylow theorems

Proof. (Sylow 2 and 3) As before, let Q be any p-subgroup of G. Now let's suppose Q is not contained in any P_i for $i = 1, 2, \ldots, r$. Thus $Q \cap P_i \leq Q$ for all i and so

$$|O_i| = |Q : Q \cap P_i| > 1 \qquad 1 \leq i \leq s$$

But this implies that $|O_i|$ is divisible by p for all i and thus $r = \sum |O_i|$ is also divisible by p, which is a constradiction as we just proved $r \equiv 1 \mod p$. Thus we must Q must be contained in some P_i, or equivalently $Q \leq gPg^{-1}$ for some $g \in G$.

Now suppose Q is a Sylow p-subgroup of G. We just showed that $Q \leq gPg^{-1}$ for some $g \in G$, but

$$|Q| = |gPg^{-1}| = p^\alpha$$

So the two subgroups must be the same. This shows that every Sylow p-subgroup is conjugate to P and thus $\mathcal{S} = Syl_p(G)$, which means

$$n_p = r \equiv 1 \mod p$$

Since all Sylow p-subgroups are conjugate to each other, we can use the orbit-stabilizer theorem to conclude

$$n_p = |G : N_G(P) \text{ for any } P \in Syl_p(G)$$

This completes the proofs of the Sylow theorems. □

Note that the second Sylow theorem shows that any two Sylow p-subgroups of a group for the same prime p are isomorphic. This leads to the following corollary

Corollary 3.9. Let P be a Sylow p-subgroup of G, the following are equivalent.

1. P is the unique Sylow p-subgroup of G, that is $n_p = 1$
2. P is normal in G: $P \trianglelefteq G$
3. All subgroups generated by elements of order p^n are p-groups. Thus if X is any subset of G such that $|x| = p^n$ for all $x \in X$, then $\langle X \rangle$ is a p-group.

Proof. $1 \leftrightarrow 2$: If there is only one p-subgroup P, then for any $g \in G$ we have

$$gPg^{-1} \in Syl_p(G) = \{P\} \rightarrow gPg^{-1} = P$$

Thus P is normal in G.

Conversely if P is normal and $Q \in Syl_p(G)$, then P and Q are conjugate by Sylow's theorem. That is there exists some $g \in G$ such that $Q = gPg^{-1} = P$ and so P is the only Sylow p-subgroup of G.

$1 \leftrightarrow 3$: Let P be the unique Sylow p-subgroup and let $X \subseteq G$ be a subset where every element $x \in X$ has prime power order: $|x| = p^\alpha$. By the second Sylow's theorem, x must be contained in some conjugate of P, thus there exists some $g \in G$ such that

$$x \in gPg^{-1} = P \rightarrow X \subseteq P \rightarrow \langle X \rangle \leq P$$

and thus $\langle X \rangle$ is a p-group.

Conversely let $\langle X \rangle$ be a p-group and suppose we choose X to be the union of all Sylow p-subgroups of G. If P is any Sylow p-subgroup, then by definition it is the p-subgroup of greatest order and since $\langle X \rangle$ is also a p-group it must be the case that $P = \langle X \rangle$ and thus be the only Sylow p-subgroup. □

A useful application of Sylow's theorems is to determine whether a group of given order can be simple.

Example. Let G be a group of order 132. We can factor this in the following ways

$$132 = 11 \cdot 12 = 3 \cdot 44 = 2^2 \cdot 33$$

noting that the prime factors are $2, 3, 11$. Using the third Sylow theorem restricts the possible values of n_p. In particular

$$\begin{aligned} n_{11} \equiv 1 \mod 11 &= \{1, 12, 23, \ldots\} \qquad & n_{11} \mid 12 \to n_{11} &= 1, 12 \\ n_3 = 1 \equiv 1 \mod 3 &= \{1, 4, 7, \ldots\} \qquad & n_3 \mid 44 \to n_3 &= 1, 4, 22 \\ n_2 \equiv 1 \mod 2 &= \{1, 3, 5, \ldots\} \qquad & n_2 \mid 33 \to n_2 &= 1, 3, 11, 33 \end{aligned}$$

Suppose G isn't simple, that is $n_p \neq 1$ for any of the prime factors p. Note that every Sylow p-group would have $p-1$ non-identity elements and so there will be $n_p(p-1)$ total elements of order p in the group G. This implies that there are

$$\begin{aligned} 120 &\text{ elements of order } 12 \\ 8 &\text{ elements of order } 3 \\ 9 &\text{ elements of order dividing } 4 \end{aligned}$$

Therefore for $n_p \neq 1$ for any prime factor p, we must have at least $120+8+9 = 137$ elements in the group, which is a clear contradiction since $|G| = 132$. Thus there must be some prime for which $n_p = 1$ and so by the corollary, there is some Sylow p-subgroup which is normal in G. Thus G is not simple.

Example. Let G be a group of order 462. We can factor this in the following ways

$$462 = 2 \cdot 3 \cdot 7 \cdot 11$$

Consider the number of Sylow 11-subgroups, which is given as

$$n_{11} \equiv 1 \mod 11 = \{1, 12, 23, \ldots\} \qquad n_{11} \mid 2 \cdot 3 \cdot 7 = 42 \to n_{11} = 1$$

Thus there is only one Sylow 11-subgroup, which means that it is normal in G by the corollary. Thus G is not simple.

Example. Consider the symmetric group S_{2p} where p is any odd prime. Note the order of the symmetric group is

$$|S_{2p}| = (2p)! = 1 \cdot 2 \cdots (p-1) \cdot p \cdot (p+1) \cdots (2p-1) \cdot 2p$$

Thus the group order has the form $p^2 m$ where m is some integer. Sylow's theorem tells us that the order of the Sylow p-subgroup is p^2, let P be that subgroup and let $x \in P$. Applying Lagrange's theorem restricts the possible order of x, specifically $|x| = 1, p, p^2$.

No elements of order p^2 exist in S_{2p} and since the identity element cannot generate any non-identity elements, we conclude that the generators of P must have order p. Therefore consider two disjoint p-cycles σ, τ and the group they generate

$$\langle \sigma, \tau \rangle = \left\{ \sigma^i \tau^j \mid 0 \leq i, j \leq p \right\}$$

There are p^2 ways to choose two integers in $[0, p)$ and so this group has order p^2, making it the desired Sylow p-subgroup, $P = \langle \sigma, \tau \rangle$. Note that P is abelian because disjoint cycles commute and powers of a cycle does not change the elements in said cycle.

3.3 The Fundamental Theorem of Finitely Generated Abelian Groups

We've already discussed quotients of groups in a previous chapter. Here we would like to easily construct larger groups from smaller ones and so give a well defined way to form group products. Using this we can classify all finite abelian groups.

Definition. *Let $G_1, G_2, \ldots$ be groups. The direct product $G_1 \times G_2 \times \cdots$ is the set of tuples (or sequences if infinite) $(g_1, g_2, \ldots)$ where $g_i \in G_i$. The group action is given by*

$$(g_1, g_2, \ldots)(h_1, h_2, \ldots) = (g_1 h_1, g_2 h_2, \ldots)$$

The identity element is $(1_{g_1}, 1_{g_2}, \ldots)$ and the group order is $|G_1||G_2| \cdots$. If any group is infinite, so is the product.

Example. Suppose $G = \mathbb{R}$, then the product $G \times G \times \cdots \times G$ is just the Euclidean n-space $\mathbb{R}^n$ with the group action being normal vector addition.

There is a useful product that we can define using the group of integers. Let $\mathbb{Z}^r = \mathbb{Z} \times \cdots \times \mathbb{Z}$ where $\mathbb{Z}^0 = 1$, this group is known as the free abelian group of rank r. Before stating the fundamental theorem of finitely generated abelian groups, we first define what it means for a group to be finitely generated.

Definition. *A group G is finitely generated if there exists a subset A of G such that $G = \langle A \rangle$.*

There is something special about finitely generated abelian groups, namely that they can be decomposed into the product of smaller groups.

Theorem 3.10 (Fundamental Theorem of Finitely Generated Abelian groups)**.**
Let G be a finitely generated abelian group, then

1. $$G \cong \mathbb{Z}_r \times \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_s}$$

   for some positive $r \in \mathbb{Z}$ and where $n_j \geq 2$, $n_{i+1} \mid n_i$, that is each subscript divides its preceding subscript and is 2 or greater.

2. This above decomposition is unique

Definition. *In the above decomposition, the free rank (or Betti number) is the value of r and the subscripts $n_1, \ldots, n_s$ are the invariant factors of the group G. The type of a group is the tuple $(n_1, \ldots, n_s)$. This is called the invariant factor decomposition of G.*

Note that a finite group must have free rank. This allows us to classify all finite abelian groups of a given order.

Corollary 3.11. The order of a finite abelian group is determined by its invariant factors.

Example. Suppose we have a group of order $n = 180 = 2^2 \cdot 3^2 \cdot 5$, then the possible invariant factors and resulting abelian groups are

Invariant Factors	Abelian Group
$2^2 \cdot 3^2 \cdot 5$	$\mathbb{Z}_{180}$
$2 \cdot 3^2 \cdot 5, 2$	$\mathbb{Z}_{90} \times \mathbb{Z}_2$
$2^2 \cdot 3 \cdot 5, 3$	$\mathbb{Z}_{60} \times \mathbb{Z}_3$
$2 \cdot 3 \cdot 5, 2 \cdot 3$	$\mathbb{Z}_{30} \times \mathbb{Z}_6$

The invariant factor decomposition does seem to be a bit random. In fact there is another way we can decompose a finite abelian group.

Theorem 3.12. Let G be an abelian group of order $n > 1$ and suppose we can decompose $n = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$, then

1. $G \cong A_1 \times \cdots A_k$, where $|A_i| = p_i^{\alpha_i}$
2. For each $A \in \{A_i\}$ where $|A| = p^\alpha$,

$$A \cong \mathbb{Z}_{p^{\beta_1}} \times \cdots \times \mathbb{Z}_{p^{\beta_t}}$$

   where $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_t \geq 1$ and $\beta_1 + \cdots + \beta_t = \alpha$
3. The above decompositions are unique

Proof. Both decomposition theorems are the consequence of a more general classification theorem pertaining to modules, which will be proved in a later chapter. □

Definition. *The integers p^{β_i} are called the elementary divisors of G and together form the elementary divisor decomposition of G.*

Example. Suppose we have a group of order $n = 360 = 2^3 \cdot 3^2 \cdot 5$, we can find the abelian groups as follows

Order p^β	Partitions of β	Abelian Groups
2^3	3; 2,1; 1,1,1	$\mathbb{Z}_8, \mathbb{Z}_4 \times \mathbb{Z}_2, \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$
3^2	2; 1,1	$\mathbb{Z}_9, \mathbb{Z}_3 \times \mathbb{Z}_3$
5	1	$\mathbb{Z}_5$

Thus all the isomorphic types are

$$\mathbb{Z}_8 \times \mathbb{Z}_9 \times \mathbb{Z}_5 \quad \mathbb{Z}_8 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_5 \quad \mathbb{Z}_4 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_9 \times \mathbb{Z}_5$$
$$\mathbb{Z}_4 \times \mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_5 \quad \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_9 \times \mathbb{Z}_5 \quad \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_5$$

It may seem a bit odd two have two different decompositions and at the same time claim both are unique. If there wasn't a way to convert between the two and thus show that they are equal we would have a very big problem on our hands. Fortunately the two *are* equal and here is how to convert between them.

Suppose we have a finite abelian group G of type $(n_1, \ldots, n_s)$, in other words

$$G \cong \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_s}$$

Suppose we define $n = n_1 \cdots n_s = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$ where we factor each $n_i = p_1^{\beta_{i1}} \cdots p_k^{\beta_{ik}}$. Then using the elementary divisor decomposition

$$\mathbb{Z}_{n_i} = \mathbb{Z}_{p_1^{\beta} i1} \times \cdots \times \mathbb{Z}_{p_k^{\beta} ik}$$

We can thus read off the elementary divisors from this, they are

$$p_j^{\beta_{ij}} \quad 1 \leq j \leq k \quad 1 \leq i \leq s \quad \beta_{ij} \neq 0$$

Example. Suppose we have a group of order $|G| = 2^3 \cdot 3^2 \cdot 5^2$ and type $(30, 30, 2)$. The invariant factors are of course $30, 30, 2$ and we can factor

$$30 = 2 \cdot 3 \cdot 5$$

Thus the elementary divisors are

$$2, 3, 5, 2, 3, 5, 2 \to 2, 2, 2 \quad 3, 3 \quad 5$$

Now for the other direction suppose we have a group of order $n = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$ and we are given the elementary divisors. We can obtain the invariant factors through the following steps

1. Group all elementary divisors which are powers of the same prime, we should obtain k lists of integers, one for each p_j
2. Arrange the integers in nonincreasing order
3. Append 1's to the end of each list to make them all the same length.
4. The i-th invariant factor is just the product of the integers in the i-th row

Example. Suppose we have a group with elementary divisors $2, 3, 2, 25, 3, 2$, we can arrange these in the following table

$p = 2$	$p = 3$	$p = 5$
2	3	25
2	3	1
2	1	1

So we can read off the invariant factors from the table. This group has type $(150, 6, 2)$ and

$$G \cong \mathbb{Z}_{150} \times \mathbb{Z}_6 \times \mathbb{Z}_2$$

3.4 Nilpotent, Solvable, and Free Groups

Definition. *A central series for a group G is a sequence of normal subgroups of G*

$$1 = G_r \trianglelefteq G_{r-1} \trianglelefteq \cdots \trianglelefteq G_2 \trianglelefteq G_1 = G$$

such that for each i we have

$$G_{i-1}/G_i \subseteq Z(G/G_i)$$

To illustrate the usefulness of central series we have the following example.

Example. Consider the Heisenberg group H, the set of 3x3 matrices of the form

$$H = \left\{ \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix} \middle| \, a, b, c \in \mathbb{R} \right\}$$

In other words it is the group of upper triangular matrices with 1s on the diagonal. We can take the following group

$$G_2 = \left\{ \begin{pmatrix} 1 & 0 & a \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \middle| \, a \in \mathbb{R} \right\}$$

and we will get the central series

$$1 = G_3 \trianglelefteq G_2 \trianglelefteq G_2 \trianglelefteq G_1 = H$$

This has some implications, namely

$$G_2/G_3 \subseteq Z(G_1/G_3) \to G_2 \subseteq Z(G_1)$$
$$G_1/G_2 \subseteq Z(G_1/G_2) \to G_1/G_2 \text{ abelian}$$

So we see that having a central series of length 2

$$1 = G_2 \trianglelefteq G_1 = G$$

will imply that G is abelian. Thus the central series gives a notion of "almost abelian" for a group, such "almost abelian" groups are called nilpotent.

Definition. *A group G is nilpotent if it has a central series of finite length.*

Note that if G is a group with $N \trianglelefteq G$, then if G/N and N are nilpotent, then G is also nilpotent.

Proposition 3.13. All finite p-groups are nilpotent.

Proof. Let G be the smallest non-nilpotent finite p-group. Note that $Z(G)$ is abelian and $G/Z(G)$ is also a p-group, thus $|G/Z(G)|$ divides $|G|$. Since p-groups have non-trivial centers, we have that $|Z(G)| > 1$ and therefore $|G/Z(G)| < |G|$. But G is the smallest non-nilpotent p-group, so $G/Z(G)$ is nilpotent by induction. Thus G must also be nilpotent and byu extension all p-groups are nilpotent. □

There is a partial converse to this, namely that for a finite group $|G| = \prod p_i^{\alpha_i}$, then it will contain a subgroup of order $p_i^{\alpha_i}$ for all i, it's Sylow p_i-subgroups.

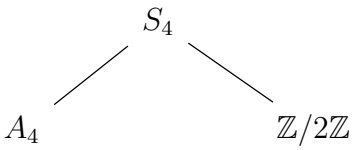
Theorem 3.14. If G is a finite nilpotent group of order $|G| = \prod p_i^{\alpha_i}$, then it has unique subgroups P_i of order $p_i^{\alpha_i}$ and

$$G = \bigoplus P_i$$

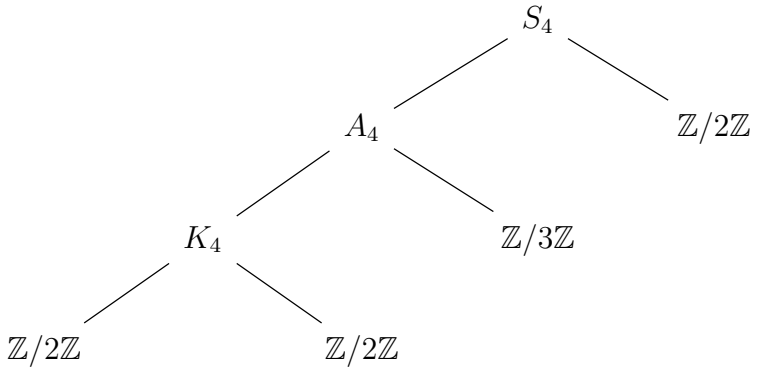
When we have a group G with a normal subgroup $N \trianglelefteq G$, we can think of G as being built from N and an extension G/N. Ideally we'd want to be able to determine G given the components $N, G/N$ but this is often not possible most of the time.

For a finite group G, we can compute $|G|$ if we know N and G/N. If N has a normal subgroup $N_1 \trianglelefteq N$, then we can further decompose N into N_1 and N/N_1 and likewise for G/N if it has a proper normal subgroup. In a sense we can continue this until we reach a group with no nontrivial proper normal subgroups which is similar to a prime factorization of groups.

Example. Take $G = S_4$ and the sign homomorphism $\epsilon : S_4 \to \pm 1$. We have the kernel $ker\epsilon = A_4$ and so G breaks up as



A cyclic group of prime order $(\mathbb{Z}/p\mathbb{Z})$ has no nontrivial subgroups and so cannot be further decomposed, however A_4 does have a normal subgroup, the Klein 4-group K_4. So we get the total decomposition



Note that factorizations of group may not be unique but all factorizations will terminate in the same multiset, that is the groups at the bottom of a factorization will always be the same.

Definition. *When all composition factors are abelian, the group G is solvable.*

These definitions allow us to classify groups into a hierarchy

$$cyclic \subset abelian \subset nilpotent \subset solvable \subset all\ groups$$

We previously introduced generators and relations as a way to describe a group. A more rigorous approach to those ideas will make use of the free group.

Definition. *The free group on generators $x_1, \ldots, x_n$ is the group of words in $x_1, \ldots, x_n$, $x_1^{-1}, \ldots, x_n^{-1}$. The elements are formed from concatenation of the generators and their inverses*

$$F(x_1, \ldots, x_n) = \left\{x_1, x_2, x_1x_2, x_1x_2^{-1}x_3^{-1}x_4^5, \ldots\right\}$$

Note that we write $x_i^2 = x_i x_i$. The group operation is concatenation

$$x_1 x_2 \cdot x_1 x_3 = x_1 x_2 x_1 x_3$$

This is sometimes called the free group of rank n.

Technically the group operation is not just concatenation but concatenation followed by reduction

$$x_i x_i^{-1} \mapsto 1$$

This will allow us to create inverses, for instance

$$\begin{aligned} x_1 x_2 \cdot x_2^{-1} x_1^{-1} &= x_1 x_2 x_2^{-1} x_1^{-1} \\ &= x_1 x_1^{-1} \\ &= 1 \end{aligned}$$

If we didn't require reduction in the group operation then inverses need not exist and instead this would be a monoid.

We will sometimes denote the free group using $F_n = F(x_1, \ldots, x_n)$. Since there is no limit to how long words can be and we can start combining arbitrarily long sequences of words, the free group can quickly become difficult to work with. We can create a group presentation using

$$\langle x_1, \ldots, x_n \mid R \rangle$$

where R is a set of elements of F_n defined to be 1, R is known as the relations of a group.

Example. Suppose we have the group presentation

$$\langle x, y \mid x^2, y^2 \rangle = \{1, x, y, xy, yx, \ldots\}$$

Formally this is a quotient of the free group $F(x, y)/G_R$ where G_R is the group generated by the elements of R.

Example. Consider the triangle group

$$\Gamma_{2,3,\infty} = \langle x, y \mid x^2.(xy)^3 \rangle$$

We can prove that this group is infinite because we have the homomorphism

$$\phi : \Gamma \to PSL_2(\mathbb{Z}) = SL_2(\mathbb{Z}) / \{\pm 1\}$$

where the generators are mapped as follows

$$x \mapsto \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad y \mapsto \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

These matrices generate $PSL_2(\mathbb{Z})$ so we see that ϕ is surjective and this Γ is infinite. In fact ϕ is an isomorphism and this triangle group provides a finite representation of $PSL_2(\mathbb{Z})$

Example. The Bunside group $B(n,m)$ is generated by $x_1, \ldots, x_n$ subject to the relations $w^m = 1$ where w is any word composed of the generators and their inverses.

We can define a free group in a different way.

Definition. *Let X be a set and F be a group, F is free on X if*

- *For every group G, there is a bijection*

$$b_G : Hom(F,G) \to Fun(X,G)$$

  where $Fun(X,G)$ is the group of functions $X \to G$ and $Hom(F,G)$ is the group of homomorphisms from $F \to G$.

- *For every homomorphism $\phi : G \to J$, the following diagram commutes*

$$\begin{array}{ccc} Hom(F,G) & \xrightarrow{b_G} & Fun(X,G) \\ \phi\downarrow & & \downarrow\phi \\ Hom(F,J) & \xrightarrow{b_J} & Fun(X,J) \end{array}$$

  That is any path with the same start and end lead to the same result.

4 Category and Representation Theory

4.1 Categories and Functors

We've previously discussed the notion of a group G acting on a set X. However X may not often be a "set." To discuss large classes of objects such as "all groups" or "all sets" we must use the right language which happens to be the language of category theory.

Definition. *A category $\mathcal{C}$ is*

- *A set of objects $Ob(\mathcal{C})$*

- *For each pair of objects $X, Y \in Ob(\mathcal{C})$, a set of morphisms denoted $Hom_{\mathcal{C}}(X,Y)$ which is the set of maps $X \to Y$*

- *Morphisms compose, if we have $f \in Hom_{\mathcal{C}}(X,Y)$ and $g \in Hom_{\mathcal{C}}(Y,Z)$ then $g \circ f : X \to Z$ is in $Hom_{\mathcal{C}}(X,Z)$. Furthermore the law is associative, if we have*

$$X \xrightarrow{f} Y \xrightarrow{g} Z \xrightarrow{h} W$$

  Then $(h \circ g) \circ f = h \circ (g \circ f) \in Hom_{\mathcal{C}}(X,W)$

- *For every $X \in Ob(\mathcal{C})$ there is an element $id_X \in Hom_{\mathcal{C}}(X,X)$, the identity morphism, which composes like an identity.*

Example. The following are some commonly found categories

- **Set**, the category of sets has all sets as objects and the morphisms are functions between sets
- **FinSet**, the category of finite sets
- **FI**, the category of finite sets where morphisms are injective functions
- **Grp**, the category of groups where the morphisms are group homomprhisms
- **Ab**, the category of abelian groups
- $\mathbf{Vect_k}$, the category of vector spaces over a field k, the morphisms are linear maps between vector spaces
- **Top**, the category of topological spaces where the morphisms are continuous maps between topological spaces.

So far morphisms are only one way, lets change this.

Definition. *A morphism $f : X \to Y$ is invertible if $\exists g : Y \to X$ such that $g \circ f = f \circ g = id_X$*

For instance an invertible morphism in the **Grp** would be the familiar isomorphism and an invertible morphism in **Top** would be a homeomorphism.

For an interesting example suppose we have a category with only one object, let's call it X. Then the set of morphisms $Hom_{\mathcal{C}}(X, X)$ is a set with an associative binary operation (composition) with an identity (id_X). This should sound familiar because this is almost like the definition for a group, in fact the set of morphisms $Hom_{\mathcal{C}}(X, X)$ in this case actually forms a monoid. If the morphisms are invertible then this monoid becomes a fully fledged group under composition. In this case although technically the group is $Hom_{\mathcal{C}}(X, X)$, we can also say that the entire category $\mathcal{C}$ is a group.

Let's formalize the category theoretic definition of a group

Definition. *A category $\mathcal{C}$ in which all morphisms are invertible is known as a groupoid. A groupoid with only a single object is a group.*

So we've defined what categories are and that's all well and good, but we now want to go places. Specifically, how can we get from one category to another.

Definition. *Let $\mathcal{C}, \mathcal{D}$ be categories. A function $F : \mathcal{C} \to \mathcal{D}$ is*

- *A function $Ob(\mathcal{C}) \to Ob(\mathcal{D})$*
- *A function $Hom_{\mathcal{C}}(X, Y) \to Hom_{\mathcal{D}}(F_x, F_y)$ such that composition still holds.*

Specifically this means that

- *If $g \circ f$ is a composition of morphisms in $\mathcal{C}$, then $F_{g \circ f} = F_g \circ F_f$ is a composition of morphisms in $\mathcal{D}$*

- $F_{id_X} = id_{F_X}$

Technically this defines a covariant functor. There are two types of functors: covariant and contravariant, with the only different being the directions of the arrows.

Definition. *A contravariant functor from $F : \mathcal{C} \to \mathcal{D}$ is the same as a covariant functor except*

- *Morphisms are mapped as $Hom_{\mathcal{C}}(X, Y) \to Hom_{\mathcal{D}}(F_y, F_x)$*
- *If $g \circ f$ is a composition of morphisms in $\mathcal{C}$, then $F_{g \circ f} = F_f \circ F_g$ is a composition of morphisms in $\mathcal{D}$*

Example. Some basic functors

- A functor of one object categories $\varphi : G_1 \to G_2$ is a group homomorphism
- We can define a functor of **FinSet** called $\oplus 2$ which sends

  $$S \mapsto S^{\oplus 2} = S \oplus S$$

  A function $f : S \to T$ is sent to $f^{\oplus 2}$ where

  $$f^{\oplus 2}(s_1, s_2) = (f(s_1), f(s_2)) \in T^{\oplus 2}$$

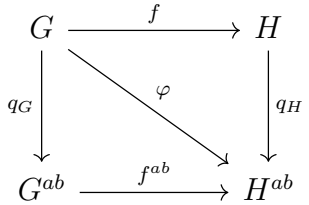
Given a an arbitrary group G, we may sometimes muse about how great life would be if G were to be abelian, alas abelian groups come once in a blue moon. However, we can actually define a functor which take a group and makes it abelian.

Example. A functor $ab : \mathbf{Grp} \to \mathbf{Ab}$ (ab for abelianization) is defined as follows. For a group G, let G' be the group generated by all commutators

$$G' = \{[x, y] \mid x, y \in G\} = \left\{xyx^{-1}y^{-1}\right\}$$

If G is abelian then it is easy to see that $[x, y] = 1$ for all $x, y \in G$ and thus G' is trivial. If G is not abelian, then G' is normal in G and we can define the quotient $G^{ab} = G/G'$. This is quotient group is abelian since we "modded out" the non abelian elements, indeed this is known as the abelianization of G.

But for this to actually be a functor we have to specify what happens to homomorphisms. Let $f : G \to H$ be a group homomorphisms and define $f^{ab} : G^{ab} \to H^{ab}$ such that the following diagram commutes



Essentially if we have any group homomorphism $\varphi : G \to H^{ab}$, then φ factors through q_G. Note that q_G, q_H are the induced quotient maps.

Now we have the example that started the whole thing, the notion of a group action.

Example. Let G be a group, considered as an one object category (let's call it $*$). We can define a functor $F : G \to \mathbf{Set}$ but first we need a function

$$F : Ob(G) \to Ob(\mathbf{Set})$$

In other words we must specify a set, let's call it X so that $F(*) = X$. We also need a function

$$F : Hom_{\mathcal{G}}(*, *) \to Hom_{\mathbf{Set}}(X, X)$$

But as we discussed earlier $Hom(*, *)$ is just our group G so really we have a function

$$F : G \to Fun(X, X)$$

Furthermore this function must be compatible with composition

$$F(g_1 g_2) = F(g_1) \circ F(g_2)$$
$$F(id_G) = id : X \to X$$

Note that these two conditions imply

$$F(gg^{-1}) = F(g) \circ F(g^{-1}) = F(id_G) = id$$

So not only is $F(g) : X \to X$ a function, it is an invertible function, a bijection from a set to itself. In other words F is a homomorphism

$$F : G \to Sym(X)$$

from a group G to the symmetric group on X. Thus a functor from a group G to **Set** is a set X together with a group action $G \curvearrowright X$.

Another type of functor allows us to condense an arbitrary category into the category of sets. In a sense this allows us to represent complex categories in terms of familiar objects, namely those of sets and functions between them.

Example. Let $\mathcal{C}$ be a category and $X \in Ob(\mathcal{C})$. We can define the functor

$$Hom_{\mathcal{C}}(X, -) : \mathcal{C} \to \mathbf{Set}$$
$$Y \mapsto Hom_{\mathcal{C}}(X, Y)$$

For a morhpism $f : Y \to Z$, we send it to the function

$$Hom_{\mathcal{C}}(X, Y) \to Hom_{\mathcal{C}}(X, Z)$$
$$\varphi \mapsto f \circ \varphi$$

Such a functor is called representable.

We also have a functor which removes any sort of structure on an object.

Definition. *The functor $L : \mathbf{Grp} \to \mathbf{Set}$ is forgetful if it sends a group G to the set of its elements.*

Such a functor essentially "forgets" any sort of group structure, turning a group back into a basic set.

Example. Let X be a set and F_X be the free group on that set. We have two functors

1. $Hom_{\mathbf{Grp}}(F_X, -)$ which sends a group $G \mapsto Hom_{\mathbf{Grp}}(F_X, G)$
2. $Hom_{\mathbf{Set}}(X, L-)$ which sends a group $G \mapsto Fun(X, G)$

So we can say that there is a bijection

$$Hom(F_X, G) \cong Fun(X, G)$$

for all groups G which is compatible with homomorphisms. This seems to imply the two functors are the same but first we need a notion of maps between functors.

Definition. *Let $\mathcal{C}, \mathcal{D}$ be categories and $F, G : \mathcal{C} \to \mathcal{D}$ be functors between the two categories. A natural transformation (or morphism between functors) is a map η which assigns to each $X \in Ob(\mathcal{C})$ a morphism $\eta_X \in Hom_{\mathcal{D}}(F_X, G_X)$ such that for each pair $X, Y \in Ob(\mathcal{C})$ and $f \in Hom_{\mathcal{C}}(X, Y)$ we have $G_f \circ \eta_X = \eta_Y \circ F_f$. In other words this diagram commutes*

$$\begin{array}{ccc} X \\ \downarrow{\scriptstyle f} \\ Y \end{array} \qquad \begin{array}{ccc} F_X & \xrightarrow{\eta_X} & G_X \\ \downarrow{\scriptstyle F_f} & & \downarrow{\scriptstyle G_f} \\ F_Y & \xrightarrow{\eta_Y} & G_Y \end{array}$$

If each η_X is an isomorphism, then η is a natural isomorphism of functors.

So we see that the functors $Hom_{\mathbf{Grp}}(F_X, -)$ and $Hom_{\mathbf{Set}}(X, L-)$ are naturally isomorphic.

A final useful concept in category theory is that of universals.

Definition. *Let $\mathcal{C}, \mathcal{D}$ be categories, $F : \mathcal{C} \to \mathcal{D}$ a functor, and $X \in Ob(\mathcal{D})$. A universal arrow from X to F is a pair $(U(X), \iota)$ where $U(X) \in Ob(\mathcal{C})$ and $\iota : X \to F_{U(X)}$ is a morphism in $\mathcal{D}$ such that if $\phi : X \to F_A$ is any morphism in $\mathcal{D}$ where $A \in Ob(\mathcal{C})$, the ϕ factors through a unique morphism F_Φ. In other words, the following diagram commutes*

$$\begin{array}{ccc} X & \xrightarrow{\iota} & F_{U(X)} \\ & \searrow^{\phi} & \downarrow{\scriptstyle F_\Phi} \\ & & F_A \end{array}$$

We will see examples of universal properties in the following chapters, namely when discussing tensor products of modules.

Example. Suppose we want a functor from a group G (considered as a category) to the category of vector spaces of a field k, $\mathbf{Vect_k}$. Let $*$ be the object in G.

If $F(*)$ is a vector space V, then for $g \in G = Hom_G(*,*)$, F_g is an invertible element of $Hom_{\mathbf{Vect_k}}(V,V)$. But an invertible map between vector transformations is just a linear transform, which can be represented with an invertible matrix so $F_g \in GL(V)$. Thus a functor from G to $\mathbf{Vect_k}$ is just a homomorphism

$$\rho : G \to GL(V)$$

This last example will be very important in following discussions.

4.2 Representations of Finite Groups

Definition. *A representation of a finite group G on a finite-dimensional vector space V is a homomorphism*

$$\rho : G \to GL(V)$$

from G to the group of automorphisms of V (ie invertible linear transformations).

When we talk about a representation, we are referring to the map + vector space pair (ρ, V) but in general if there is no ambiguity we will say that V is a representation of G. In this same vein if $v \in V$ is a vector then we will write gv instead of $\rho(g)(v)$. The dimension of a representation is the associated dimension of the vector space.

Definition. *A homomorphism of representations $(\rho, V) \to (\psi, W)$ is a linear map $f : V \to W$ such that*

$$\psi(g) \circ f = f \circ \rho(g) \quad \to \quad f(gv) = gf(v)$$

or in other words the following diagram commutes for all $g \in G$

$$\begin{array}{ccc} V & \xrightarrow{\quad f \quad} & W \\ {\scriptstyle g}\downarrow & & \downarrow{\scriptstyle g} \\ V & \xrightarrow{\quad f \quad} & W \end{array}$$

Example. We can represent the group $\mathbb{Z}/n\mathbb{Z}$ on $\mathbb{R}^2$ by rotating counterclockwise $2\pi/n$

$$\rho : \mathbb{Z}/n\mathbb{Z} \to GL(\mathbb{R}^2) = GL_2(\mathbb{R}) \qquad \rho(1) = \begin{pmatrix} \cos(2\pi/n) & -\sin(2\pi/n) \\ \sin(2\pi/n) & \cos(2\pi/n) \end{pmatrix}$$

This also suggests a complex representation

$$\rho : \mathbb{Z}/n\mathbb{Z} \to GL(\mathbb{C}) \qquad \rho(1) = e^{2\pi i/n}$$

Note that $GL(\mathbb{C})$ is just the multiplicative group of invertible complex numbers (ie not 0). In general we have $|\rho(z)| = 1$, that is they are roots of unity. If we choose to set $\rho(z) = 1$ for all $z \in \mathbb{Z}/n\mathbb{Z}$, then we obtain the trivial representation.

Example. The symmetric group S_3 has a representation on $\mathbb{R}^3$ by permuting coordinates, for instance

$$\rho((1\ 2)) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \rho((1\ 2\ 3)) = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Not that we can also represent S_3 using a subspace $V \subseteq \mathbb{R}^3$ given by

$$V = \{(x, y, z) \mid x + y + z = 0\}$$

since permuting the coordinates will not change the sum.

This last example gives a notion of representations containing other representations which we will explore further

Definition. *A subrepresentation of V is a subspace $W \subset V$ which is invariant under G. If the only subrepresentations of V are V and 0, then V is an irreducible representation (irrep).*

Proposition 4.1. If W, W' are subrepresentations of V with $W \cap W' = 0$ (disjoint) and $W + W' = V$, then

$$V = W \oplus W'$$

In other words, representations and subrepresentations behave just like we expect vector spaces and their subspaces to behave, at least when talking about direct sums. We can use this idea later to build up representations from "smaller" representations.

Theorem 4.2. Let G be a finite group, V a representation of G and $W \subset V$ a subrepresentation. Then there exists a complementary subrepresetation $W' \subset V$ such that $V = W \oplus W'$.

Proof. Let $W_0 \subset V$ be a complementary subspace, that is

$$W \cap W_0 = 0 \qquad W + W_0 = V$$

Define the projection map $\rho_0 : V \to W$ so that $\ker \rho_0 \subseteq W_0$. This means that every vector $v \in V$ can be written as

$$v = w + w_0 \quad \to \quad \rho_0(v) = \rho_0(w + w_0) = w$$

Now define the map $\rho : V \to W$ given by

$$\rho = \frac{1}{|G|} \sum_{g \in G} g \circ \rho_0 \circ g^{-1}$$

Let $W' = \ker \rho$ and consider some vector $w' \in W'$. For some element $h \in G$

$$\rho(hw') = \frac{1}{|G|} \sum_{g \in G} g\rho_0 g^{-1}(hw')$$

$$h^{-1}\rho(hw') = \frac{1}{|G|} \sum_{g \in G} (h^{-1}g)\rho_0(h^{-1}g)^{-1}w'$$

But $h, g \in G$ so we can just redefine the summands since they are elements of the group

$$h^{-1}\rho(hw') = \frac{1}{|G|}\sum_{g\in G} g\rho_0 g^{-1}w' = \rho(w') = 0$$

Thus W' preserves the group action (because $gw' \in W'$) and so $W' \subset V$ is the desired complementary subrepresentation. □

This theorem has an important consequence.

Corollary 4.3. Any representation is a direct sum of irreducible representations.

This property is sometimes referred to as complete reducibility or semisimplicity.

Example. Suppose we have a representation of $\mathbb{Z}$ on $\mathbb{C}^2$ given by

$$\rho : \mathbb{Z} \to GL_2(\mathbb{C}) \qquad \rho(1) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

Let's define the subspace

$$W = \mathbb{C}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \left\{ \begin{pmatrix} z \\ 0 \end{pmatrix} \middle| \, z \in \mathbb{C} \right\} \qquad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} z \\ 0 \end{pmatrix} = \begin{pmatrix} z \\ 0 \end{pmatrix}$$

So we see that W is a copy of the trivial representation and for there to be a complementary subrepresentation W' it would have to $1D$. In other words it is spanned by an eigenvector of $\rho(1)$

Example. Let $G = \mathbb{Z}/n\mathbb{Z}$, then every irreducible representation of G are 1D and of the form

$$\rho : \mathbb{Z}/n\mathbb{Z} \to GL(\mathbb{C}) = \mathbb{C}^* \qquad \rho(1) = e^{2\pi i k/n}$$

To see this let V be an irreducible representation of $\mathbb{Z}/n\mathbb{Z}$ and define $\gamma = \rho(1) \in GL(V)$. Let v be an eigenvector of γ so that

$$\gamma v = \lambda v \to \gamma^m v = \lambda^m v$$

for all powers of γ. This implies that the action preserves $\mathbb{C}v$, but V is irreducible so $\mathbb{C}v = V$. In other words V is generated by a single vector v, so it is one dimensional.

Now that we've established that we can decompose every representation into a direct sum of irreducible ones, we can ask the question of uniqueness. This question is answered by the next lemma.

Lemma 4.4 (Schur)**.** Let V, W be irreducible representations of G and suppose $\varphi : V \to W$ is a homomorphisms of representations, then

1. Either $\varphi = 0$ or it is an isomorphism
2. If it is an isomorphism, then φ is a scalar $\varphi = \lambda \cdot I$ where I is the identity matrix.

Proof. Since $\ker\varphi$ and $\text{im }\varphi$ are invariant subspaces of V and W respectively, for them to be irreducible it must be the case that

$$\ker\varphi = 0 \qquad \text{im }\varphi = W$$

Which would imply that φ is an isomorphism

The map φ has some eigenvalue λ, which implies $\varphi - \lambda I$ is a map with nontrivial kernel. But this means that it must be the zero map so

$$\varphi - \lambda I = 0 \to \varphi = \lambda I$$

Thus φ is a scalar map. □

Example. Suppose we have a finite abelian group G and let V be a complex representation of G. Note that in general for some $g \in G$, the automorphism $\rho(g) : V \to V$ is not linear because

$$g(hv) \neq h(gv)$$

However if g is in the center $Z(g)$, then this is linear for any ρ. In particular if V is irreducible, then g acts as a scalar map. But this means that every subspace of V is invariant and so it must be that V is one dimensional. The irreducible representations of G are thus homomorphisms

$$\rho : G \to \mathbb{C}^*$$

Example. Now lets consider a slightly more involved example. Suppose we want to find the irreducible complex representations of S_3, the simplest non-abelian group. There are two obvious 1D representations, the trivial one U and the alternating representation U'.

$$\begin{aligned} U : \sigma v = v \qquad & \rho(\sigma) = 1 \\ U' : \sigma v = sgn(\sigma)v \qquad & \rho(\sigma) = sgn(\sigma) = \pm 1 \end{aligned}$$

where $\sigma \in S_3$ and $v \in \mathbb{C}$. We also have the action of S_3 on $\mathbb{C}^3$ by coordinate permutation, called the standard representation. This action leaves the space spanned by $(1,1,1)$ invariant, which suggest the existence of a complementary subspace. Let $W' = span(1,1,1)$, we can express any vector $v \in \mathbb{C}^3$ as a sum $v = w + w'$ where W is the complementary subspace. We find that

$$W = \{(x,y,z) \mid x + y + z = 0\}$$

We will find that W is the 2D irreducible representation of S_3, but for now we would like to verify this. Consider some arbitrary representation V and consider some 3-cycle τ (for instance $\tau = (1\,2\,3)$). Since ρ is a homomorphism

$$\rho : S_3 \to GL(V) \qquad \rho(\tau^3) = \rho^3(\tau) = 1$$

In other words, this is a linear transformation of order 3 and so it will have some eigenvalues $\lambda = 1, \omega, \omega^2$ where $\omega = 2\pi i/3$ is the 3rd root of unity. Thus V will be spanned by the corresponding eigenvectors and we can write

$$V = V_1 \oplus V_\omega \oplus V_{\omega^2}$$

Now consider the action of some transposition σ (for instance $\sigma = (1\,2)$) on V. Note that σ and τ will generate all of S_3. If we take any vector v then

$$\begin{aligned}\tau\sigma v &= \sigma\tau^2 v \\ &= \sigma\omega^{2i} v \\ &= \omega^{2i}\sigma v\end{aligned}$$

So we see that σv is also an eigenvector, indeed if v is an eigenvector of τ with eigenvalue ω^i, then σv is an eigenvector with eigenvalue ω^{2i}.

Suppose we have an arbitrary eigenvector v of τ. If it's eigenvalue is not 1, then $\omega^{2i} \neq \omega^i$ and so $v, \sigma v$ are linearly independent vectors and will span a 2D subspace W of V. If we take $\tau = (1\,2\,3)$ and $\sigma = (1\,2)$ we can see that this subspace is isomorphic to the standard representation derivative earlier and thus it is irreducible.

Now if v has an eigenvalue of 1, then $v, \sigma v$ may or may not be linearly independent. If they aren't, then we can take $v_1 = v + \sigma v, v_2 = v - \sigma v$ which are linearly independent. Either way, one vector will span a space isomorphic to the trivial representation and the other will span a space isomorphic to the alternating representation. Thus

$$V = U^{\oplus a} \oplus U'^{\oplus b} \oplus W^{\oplus c}$$

The multiplicities of each irreducible representation can be found using the multiplicities of each eigenvalue. c is the number of independent eigenvectors with eigenvalue ω. $a + c$ is the multiplicity of 1 as an eigenvalue of σ, $b + c$ is the multiplicity of -1 as an eigenvalue of σ.

4.3 Character Theory

As the last example in the previous demonstrated, trying to describe arbitrary representations is tedious and hard and should be avoided if not for the purpose of intellectual enrichment. For a general group it's easier to work in terms of characters, a function which condenses essential information about a representation.

Definition. *Let V be a complex representation of some group G. The character if a function*

$$\chi_V : G \to \mathbb{C} \qquad \chi_V(g) = Tr(\rho(g))$$

For a general representation over some field k, the character is a function

$$\chi_v : G \to k$$

First let's examine some properties of this function.

$$\begin{aligned}\chi_v(hgh^{-1}) &= Tr(\rho(hgh^{-1})) \\ &= Tr[\rho(h)\rho(g)\rho(h^{-1})] \\ &= Tr(\rho(g)) \\ &= \chi_V(g) \\ \chi_V(1) &= Tr(\rho(1)) \\ &= Tr(id_{GL(V)}) \\ &= \dim V\end{aligned}$$

The first calculation indicates that the character is what's known as a class function, it is constant on the conjugacy classes of G. The second gives us a useful way to find dimensions of representations.

Example. Consider the representation of S_3 on $\mathbb{C}^3$, we have the following characters

$$\chi_V(id) = Tr(I) = 3 \qquad \chi_V((1\,2)) = 1 \qquad \chi_V((1\,2\,3)) = 0$$

There's a pattern that we can see from this exercise.

Proposition 4.5. If G acts on a finite set X, then there is a permutation representation of G on $\mathbb{C}^{|X|}$ where G permutes basis elements as it permutes X. Furthermore

$$\chi_{\mathbb{C}^{|X|}} = \text{number of elements of } X \text{ fixed by } g$$

Characters also behave as we expect under direct sums. Suppose $V = V_1 \oplus V_2$, then we can interpret the matrix as

$$\rho_V(g) = \left(\begin{array}{c|c} \rho_{V_1}(g) & \\ \hline & \rho_{V_2}(g) \end{array}\right)$$

Where $\rho_{V_1}(g)$ and $\rho_{V_2}(g)$ are square matrices, we thus see that

$$Tr(\rho_V(g)) = Tr(\rho_{V_1}(g)) + Tr(\rho_{V_2}(g))$$
$$\therefore \chi_V = \chi_{V_1} + \chi_{V_2}$$

We can express the information about irreducible representations of a group G in the form a character table. The top row lists all the conjugacy classes of G along with a representative g and the number of elements in each class. Each row lists a irreducible representation along with its character on each conjugacy class.

Example. Consider again the complex representations of S_3. The conjugacy classes are the identity [1], transpositions $[(1\,2)]$ and 3-cycles, $[(1\,2\,3)]$. It's a simple exercise to check the sizes of these conjugacy classes. The trivial representation will take on values $(1, 1, 1)$ and the alternating representation takes on $(1, -1, 1)$, these are straightforward. For the standard representation, note that the permutation representation decomposes as

$$\mathbb{C}^3 = U \oplus W$$

where U is the trivial representation and W is the standard representation. The permutation representation has character $(3, 1, 0)$ which we derived in the previous example. Thus we can use this to write down the character of the standard representation and we get the character table of S_3

	1	3	2
	1	$(1\,2)$	$(1\,2\,3)$
trivial U	1	1	1
alternating U'	1	-1	1
standard W	2	0	-1

This table also solves the problem in the previous section of determining multiplicities of irreducible constituents of some arbitrary representation. In particular if

$$V \cong U^{\oplus a} \oplus U'^{\oplus b} \oplus W^{\oplus c}$$

then the characters will satisfy

$$\chi_V = a\chi_U + b\chi_{U'} + c\chi_W$$

The characters are linearly independent so we can determine any representation V up to isomorphism using its character χ_V.

Example. Consider a complex representation of $\mathbb{Z}/3\mathbb{Z}$ and let $\omega = \exp(2\pi i/3)$. The character table is

	1	1	1
	0	1	2
$triv$	1	1	1
ω	1	ω	ω^2
ω^2	1	ω^2	ω

4.4 The Projection Formula and Orthogonality

In the previous sections we showed how to decompose a representation into irreducible ones, but we still don't know what happens to each vector in the original representation. Here, we will discuss the projection of a vector in the trivial part of its decomposition.

For a representation V of a group G, consider

$$V^G = \{v \in V \mid gv = v \quad \forall g \in G\}$$

In other words this is the set of all vectors on which every g acts trivially. We want an explicit way to find V^G and in fact we have a projection formula

$$\phi = \frac{1}{|G|} \sum_{g \in G} g$$

It can be shown that the map ϕ is a projection $V \to V^G$. The dimension of V_G is the number of times the trivial representation shows up in a decomposition, we can find this number

$$\begin{aligned} \dim V^G &= Tr(\phi) \\ &= \frac{1}{|G|} \sum_{g \in G} Tr(g) \\ &= \frac{1}{|G|} \sum_{g \in G} \chi_V(g) \end{aligned}$$

This seems to imply that for a non-trivial irreducible representation, the sum of the character $\chi_V(g)$ over all elements $g \in G$ is zero.

We can explore this idea further. Consider the set of homomorphisms between vector spaces (linear maps) which are compatible with G, $Hom_G(V, W)$ (these are actually G-module homomorphisms but that's a later discussion). If V, W are irreducible then by Schur's theorem we have

$$\dim Hom_G(V, W) = \begin{cases} 1 & \text{if } V \cong W \\ 0 & \text{if } V \ncong W \end{cases}$$

We can express the space of homomorphisms as $Hom_G(V, W) = V^* \otimes W$ where V^* is the dual space. Note that

$$\chi_{V^*} = \overline{\chi_V} \qquad \chi_{V \otimes W} = \chi_V \chi_W$$

Thus we get the formula

$$\chi_{Hom} = \overline{\chi_V} \chi W$$

Combining the previous two facts along with our discussion of projections we get

$$\dim Hom_G(V, W) = \frac{1}{|G|} \sum_{g \in G} \overline{\chi_V}(g) \chi_W(g) = \begin{cases} 1 & \text{if } V \cong W \\ 0 & \text{if } V \ncong W \end{cases}$$

Theorem 4.6. Define the Hermitian inner product

$$\langle \chi_{V_i}, \chi_{V_j} \rangle = \frac{1}{|G|} \sum_{g \in G} \overline{\chi_{V_i}(g)} \chi_{V_j}(g)$$

The characters of irreducible representations are orthonormal under this inner product, that is

$$\langle \chi_{V_i}, \chi_{V_j} \rangle = \delta_{ij}$$

where δ is the Kronecker delta.

Corollary 4.7. The fact that characters are orthonormal has major consequences:

- The number of irreducible representations of G is less than or equal to the number of conjugacy classes of G.
- A representation is uniquely determined by its character up to isomorphism. In other words if V, V' have the same character then they are isomorphic.
- A representation V is irreducible if and only if

$$\langle \chi_V, \chi_V \rangle = 1$$

- The multiplicity a_i of some V_i in V is the inner product

$$a_i = \langle \chi_V, \chi_{V_i} \rangle$$

Regarding the first point, it can be shown using other projection formulas that the number of irreducible representations is equal to the number of conjugacy classes.

Example. Consider a representation of S_3 by V where V is the space of homogeneous degree 3 polynomials in 3 variables, that is

$$V = span(x^3, y^3, z^3, x^2y, x^2z, y^2z, xy^2, xz^2, yz^2, xyz)$$

S_3 acts by variable permutations on V and we can see that

$$\begin{aligned}\chi_\sigma = Tr(\rho(\sigma)) &= \text{number of 1s on diagonal} \\ &= \text{number of fixed basis elements}\end{aligned}$$

Note that id fixes everything, $(x\,y)$ fixes the subspace spanned by $\{z^3, zyx\}$, and $(x\,y\,z)$ fixes the subspace spanned by $\{xyz\}$. Thus

$$\begin{aligned}\langle \chi_V, \chi_U \rangle &= \frac{1}{6}\sum_\sigma \chi_V(\sigma)\chi_U(\sigma) \\ &= \frac{1}{6}\left[\overline{\chi_V(id)}\chi_U(id) + 3\overline{\chi_V(x\,y)}\chi_U(x\,y) + 2\overline{\chi_V(x\,y\,z)}\chi_U((x\,y\,z))\right] \\ &= \frac{1}{6}(10\cdot 1 + 3\cdot 2\cdot 1 + 2\cdot 1\cdot 1) = 3 \\ \langle \chi_V, \chi_{U'} \rangle &= \frac{1}{6}\sum_\sigma \chi_{U'}(\sigma)\chi_U(\sigma) \\ &= \frac{1}{6}(10\cdot 1 + 3\cdot 2\cdot (-1) + 2\cdot 1\cdot 1) = 1 \\ \langle \chi_V, \chi_W \rangle &= \frac{1}{6}\sum_\sigma \chi_W(\sigma)\chi_U(\sigma) \\ &= \frac{1}{6}(10\cdot 2 + 3\cdot 2\cdot 0 + 2\cdot 1\cdot (-1)) = 3\end{aligned}$$

and so this representation breaks up as

$$V \cong U^{\oplus 3} \oplus U' \oplus W^{\oplus 3}$$

Example. The group $\mathbb{Z}/n\mathbb{Z}$ has n conjugacy classes and thus has n irreducible representations, each one corresponds to a n-th root of unity.

Definition. *Let G be a finite group which acts on itself by left multiplication. The induced permutation representation V_{reg} is called the regular representation.*

Recall that for a permutation representation, the character corresponds to the number of elements fixed by the group action. But left multiplication does not fix any element except the identity, so

$$\chi_{reg}(g) = \begin{cases} 0 & g \neq 1 \\ |G| & g = 1 \end{cases}$$

Now lets consider irreducible constituents of the regular representation. Suppose V_i is a irreducible representation of G, then

$$\begin{aligned}\langle \chi_{reg}, \chi_{V_i} \rangle &= \frac{1}{|G|} \sum_{g \in G} \overline{\chi_{reg}(g)} \chi_{V_i}(g) \\ &= \frac{1}{|G|} \overline{\chi_{reg}(1)} \chi_{V_i}(1) \\ &= \dim V_i\end{aligned}$$

Since $\dim V_i > 0$, every irreducible representation is a direct summand of the regular representation. Furthermore this implies

$$\dim V_{reg} = \dim \bigoplus V_i^{\dim V_i} \to |G| = \sum_{\text{irreps}} (\dim V_i)^2$$

We can confirm this for the case of S_3, there are two 1D and one 2D irreducible representation and it is indeed the case that

$$|S_3| = 1^2 + 1^2 + 2^2 = 6$$

Example. Let's put together everything we've learned about representations to derive the character table of S_4. There are six conjugacy classes corresponding to the six ways we can partition 4. These classes are given by the representatives $id, (1\,2), (1\,2\,3), (1\,2\,3\,4), (1\,2)(3\,4)$. It's simple to compute the sizes of each class, they are $1, 6, 8, 6, 3$ respectively.

The first few irreducible representations are simple to see, we have two 1D irreducible representations, the trivial U and alternating U'. The character of U is $(1, 1, 1, 1, 1)$ as required and we can calculated the sign of each conjugacy representation to get the character of U' as $(1, -1, 1, -1, 1)$. Furthermore there is a permutation representation of S_4 on $\mathbb{C}^4$ with character $(4, 2, 1, 0, 0)$. Just like the case of S_3, this fixes the space spanned by $(1, 1, 1, 1, 1)$ and so the character will break up as $\chi_W + \chi_U$. Thus we can derive the standard representation W with character $(3, 1, 0, -1, -1)$.

So far we have two 1D irreducible representations and one 3D irreducible representation, the sum of square dimensions is $1 + 1 + 9 = 11$, which is definitely not 24. We expect there to be two more representations (for a total of 5 since there are 5 conjugacy classes) with square dimensions summing to 13. The only possible case for this is if we have one 2D representation and one 3D representation $(4 + 9 = 13)$.

We can obtain the 3D representation by tensoring the standard representation W with the alternating representation U'. The resulting representation $V' = W \otimes U'$ has character $(3, -1, 0, 1, -1)$ which is orthogonal to the other representations. This representation is given by

$$\rho_{W \otimes U'}(\sigma) = \rho_W(\sigma)\epsilon(\sigma)$$

where ϵ is the sign homomorphism. We also see that $\langle \chi_{V'}, \chi_{V'} \rangle = 1$ so this is indeed the irreducible representation we are looking for. The last irreducible representation we can derive from orthogonality, its character is $(2, 0, -1, 0, 2)$.

Note that the character of the conjugacy class $[(1\,2)(3\,4)]$ is equal to the character of the identity. This implies that this conjugacy class acts like the identity and this is a representation of

$$S^4/\{id, (1\,2)(3\,4), (1\,3)(2\,4), (1\,4)(2\,3)\} \cong S_3$$

In other words this just the standard representation of S_3 pulled back to S_4, we call this the inflated representation from S_3. The full character table is thus

	1	6	8	6	3
	id	$(1\,2)$	$(1\,2\,3)$	$(1\,2\,3\,4)$	$(1\,2)(3\,4)$
trivial U	1	1	1	1	1
alternating U'	1	-1	1	-1	1
standard W	3	1	0	-1	-1
$V' = W \otimes U'$	3	-1	0	1	-1
inflated W'	2	0	-1	0	2

Part III
Ring and Module Theory

5 Introduction to Rings

5.1 Definition and Properties

So far we've been studying groups, sets endowed with a single binary operation. For instance we can consider the group of integers or the group of real numbers. But integers and real numbers have more operations than just addition, for instance we can multiply them as well. To study these sets we must consider structures with two binary operations.

Definition. *A ring is a set R with two binary operations $+, \cdot$ such that*

1. *$(R, +)$ is an abelian group*
2. *$\cdot$ is associative*
$$(a \cdot b) \cdot c = a \cdot (b \cdot c)$$
3. *$\cdot$ and $+$ distribute as*
$$a \cdot (b + c) = a \cdot b + a \cdot c$$
$$(a + b) \cdot c = a \cdot c + b \cdot c$$

Just like normal multiplication we will usually write $a \cdot b = ab$. The additive identity will be denoted with 0 and the additive inverse of x as $-x$. Just like with groups, there are some special subsets of rings.

Definition. *A ring R in which $\cdot$ commutes, that is $ab = ba$ is a commutative ring.*

Generally we will study commutative rings but some theorems will also hold in the case of a noncommutative ring.

Definition. *A ring with identity is a ring R with a multiplicative identity, usually denoted 1. This element will satisfy*
$$1 \cdot x = x \cdot 1 = 1$$

Definition. *A ring R with identity 1, $1 \neq 0$, is a division ring if for every element $x \neq 0$ there is another element $b \in R$ such that $ab = ba = 1$. In other words multiplicative inverses exist for every nonzero element. A commutative division ring is called a field*

Example. The integers $(\mathbb{Z}, +, \cdot)$ is a commutative ring with identity but is not a field. The rings $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are all fields.

$(\mathbb{Z}/n\mathbb{Z}, +, \cdot)$ is a commutative ring with identity, if n is prime then this is a finite field.

Example. The real Hamilton Quaternions

$$\mathbb{H} = \left\{a + bi + cj + dk \mid a, b, c, d \in \mathbb{R}, i^2 = j^2 = k^2 = -1\right\}$$

form a division ring but not a field (not commutative). To see that every element is invertible suppose we define the function

$$Norm(a + bi + cj + dk) = a^2 + b^2 + c^2 + d^2$$

we see then that

$$Norm(x) = x\overline{x} = (a + bi + cj + dk)(a - bi - cj - dk)$$

thus

$$\frac{x\overline{x}}{Norm(x)} = 1 \to x^{-1} = \frac{\overline{x}}{Norm(x)}$$

This implies that an element is invertible if $Norm(x) \neq 0$, but for $Norm(x) = 0$ would require that $a^2 + b^2 + c^2 + d^2 = 0$ which forces $a = b = c = d = 0$ which is the zero element. Thus every nonzero element is invertible.

Proposition 5.1. Some simple properties of ring elements

- $a \cdot 0 = 0 \cdot a = 0$
- $(-a) \cdot b = a \cdot (-b) = -(a \cdot b)$
- $(-a) \cdot (-b) = a \cdot b$
- If $1 \neq 0$, then $(-1) \cdot a = a \cdot (-1) = -a$

These properties show that the additive and multiplicative parts of a ring behave exactly like we'd expect them to.

Definition. *A zero divisor is an element $x \in R$ such that there exists another $y \in R$, $x, y \neq 0$ and $xy = 0$ or $yx = 0$.*

Definition. *Let R be a ring with $1 \neq 0$, an element $u \in R$ is called a unit if it has a multiplicative inverse, that is there exists $v \in R$ such that $uv = vu = 1$.*

Proposition 5.2. An element cannot be both a unit and a zero divisor

Proof. Suppose $a \in R$ is both a unit and a zero divisor with $c \in R$ such that $ac = 1$ and $b \in R$, $b \neq 0$ such that $ab = 0$ or $ba = 0$. In either case we have

$$\begin{aligned} ab = 0 &\to cab = c0 = 0 \\ &\to 1b = b = 0 \\ ba = 0 &\to bac = 0c = 0 \\ &\to b1 = b = 0 \end{aligned}$$

but $b \neq 0$ so this is a contradiction. Thus a cannot be both a unit and a zero divisor $\square$

Definition. *The set of units, denoted*

$$R^\times = \{u \in R \mid u \text{ is a unit}\}$$

forms a group under multiplication.

Definition. *A commutative ring with* $1 \neq 0$ *and no zero divisors is an integral domain.*

Example. $\mathbb{Z}$ is an integral domain

$\mathbb{Z}/n\mathbb{Z}$ has zero divisors when n is composite, for instance when $n = 12$ we have $3 * 4 = 0$

Fields have no zero divisors so they are all integral domains

Proposition 5.3. Suppose $a \in R$ is not a zero divisor. If $ab = ac$ then either $a = 0$ or $b = c$

Proof.

$$ab = ac \to ab - ac = 0 \to a(b - c) = 0$$

Since a is not a zero divisor, we must have $a = 0$ or $b - c = 0 \to b = c$ □

This proposition allows us to use familiar laws of cancellation. Using this, we can prove a partial converse to the last example.

Proposition 5.4. A finite integral domain is a field

Proof. Let R be a finite integral domain and consider some $a \in R$. Consider a map $x \mapsto ax$ from $R \to R$. This map is injective by the cancellation law which also means it is surjective because R is finite. This means that there must be an element mapping to the multiplicative identity, that is there exists a $b \in R$ such that $ab = 1$. We can do this for any $a \in R$, demonstrating that all nonzero elements are units, thus R is a field. □

Just like how we can define subgroups of larger groups, we can also define subrings.

Definition. *A subring is a set* $S \subseteq R$ *such that*

1. $(S, +)$ *is a subgroup of* $(R, +)$ *(closed under addition)*
2. $\forall a, b \in S$, $ab \in S$ *(closed under multiplication)*

In other words S *by itself is a ring.*

Also like what we did for groups, there is a simple subring test.

Proposition 5.5. For $S \subseteq R$, if

1. $S \neq \varnothing$
2. $\forall a, b \in S$, $a - b \in S$
3. $\forall a, b \in S$, $ab \in S$

then S is a subring of R.

There a few observations to be made here. If R has an identity 1, it is not necessary that S also have that identity. While the subring of an integral domain is also an integral domain, the subring of a field may not be a field.

Example. $\mathbb{Z} \subseteq \mathbb{Q}$ is a subring so $\mathbb{Z}$ is an integral domain but yet it is not a field.

Example. Consider the Gauss integers

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

This is a subring of $\mathbb{C}$. Furthermore this is an integral domain with units

$$(\mathbb{Z}[i])^\times = \{\pm 1, \pm i\}$$

5.2 More Examples

Here we will introduce some useful examples which will show up frequently in following sections.

Definition. *Let R be a commutative ring with identity, then the ring of polynomials in the variable x with coefficients in R is the set of formal sums*

$$R[x] = \left\{ \sum_{i=0}^{n} a_i x^i \,\middle|\, a_i \in R \right\}$$

For a polynomial $f \in R[x]$, if $a_n \neq 0$ then $deg(f) = 0$ is its degree.

We can also define polynomials in more than 1 variable, for instance

$$R[x, y] = \left\{ \sum_{i,j} a_{ij} x^i y^j \right\}$$

The ring operations are addition and multiplication which behaves just like we expect from standard algebra. Addition is component-wise and multiplication is accomplish through expanding and collecting like terms. Note that the ring R is a subring of $R[x]$ as the ring of constant polynomials.

Proposition 5.6. Let R be an integral domain

1. For polynomials $f, g \in R[x]$, $deg(fg) = deg(f) = deg(g)$
2. The units of $R[x]$ are the units of R, that is $(R[x])^\times = R^\times$
3. $R[x]$ is also an integral domain

Proof. Consider two nonzero polynomials $f, g \in R[x]$ and suppose

$$f = \sum i = 0^m a_i x^i \quad g = \sum_{j=0}^{n} b_j x^j$$

it's straightforward to see then that

$$fg = a_m b_n x^{n+m} + \cdots$$

and since R is an integral domain, $a_m b_n \neq 0$ so $deg(fg) = m + n$ as required. This also shows that in general if $f, g \neq 0$ then there will at least be one nonzero term in the product, so zero divisors cannot exist which proves (3).

Suppose we have a unit $a \in R^\times$ and $ab = 1$. We can interpret these elements as constant polynomials in $R[x]$ and so $a \in (R[x])^\times$, thus showing that $R^\times \subseteq (R[x])^\times$. For the other direction, suppose that $f, g \in R[x]$ are polynomials such that $fg = 1$. By the previously proved proposition this implies

$$deg(fg) = deg(f) + deg(g) = 0 \rightarrow deg(f) = deg(g) = 0$$

In other words they are constant polynomials which multiply to 1 and so they will be units in R, thus $(R[x])^\times \subseteq R^\times$. Therefore the two sets are equal. □

Some other properties of rings also carry over to polynomial rings. If R has zero divisors, then so will $R[x]$. If $f \in R[x]$ is a zero divisor, then there will exist some $c \in R$ such that $cf = 0$. If $S \subset R$ is a subring, then $S[x] \subset R[x]$ is also a subring.

Definition. *Let R be a ring with $1 \neq 0$. $M_n(R)$ is the ring of $n \times n$ matrices with entries from R.*

This is a ring under usual matrix addition and multiplication so for $n \geq 2$, the ring will not be commutative. Obviously if $n = 1$ we just have a ring of scalars which is just R. The multiplicative identity is of course the identity matrix and the units form the group of invertible matrices $GL_n(R)$. Just like with polynomial rings if $S \subseteq R$ is a subring, then $M_n(S)$ is a subring of $M_n(R)$.

Definition. *Let R be a commutative ring with $1 \neq 0$ and G be any finite group. We will denote the group operation with concatenation. The group ring is the set of formal sums*

$$RG = \left\{ \sum_i a_i g_i \middle| a_i \in R, g_i \in G \right\}$$

Addition in this ring is component-wise and multiplication is performed by defining

$$(ag_i)(bg_j) = (ab)(g_i g_j) = abg_k \qquad g_i g_j = g_k$$

and following standard distributive laws. Note that RG is a commutative ring if G is abelian. By convention we will require that

$$a1_G = a \quad 11_G = 1 \quad 1g = g$$

Example. Consider the group ring $R = \mathbb{Z}D_8$ and two elements $\alpha = 1 - 3r^2, \beta = 2s - sr^2$. Then

$$\begin{aligned}\alpha + \beta &= 1 + 2s - r^2 - sr^2 \\ \alpha\beta &= (1 - 3r^2)(2s - sr^2) \\ &= 2s - sr^2 - 6r^2s + 3r^2sr^2 \\ &= 2s - sr^2 - 6sr^2 + 3s \\ &= 5s - 7sr^2\end{aligned}$$

Note that both the ring R and group G shows up in the group ring. R is just the subring of scalars, for any elements $a \in R$, we have $a \cdot 1_G \in RG$. Similarly, G is the set of units of RG. If G is a finite group, then RG will have zero divisors.

5.3 Homomorphisms and Quotient Rings

Just like maps between groups, we can construct maps between rings which preserve the ring structure.

Definition. *A ring homomorphism φ from R to S is a group homomorphism*

$$\varphi : (R, +) \to (S, +)$$

which also satisfies

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

By convention we will require that $\varphi(0) = 0$, $\varphi(1) = 1$, though this is not necessary. As usual a bijective ring homomorphism is a ring isomorphism. We can also define the kernel and image of a ring homomorphism in the usual way

$$\begin{aligned}\ker \varphi &= \{a \in R \mid \varphi(a) = 0_S\} \\ \text{im } \varphi &= \{\varphi(a) \mid a \in R\}\end{aligned}$$

Example. We can define a ring homomorphism

$$\pi : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z} \qquad \pi(a) = \overline{a}$$

This satisfies all the properties due to how modulo behaves, furthermore

$$\ker \pi = n\mathbb{Z} \qquad \text{im } \pi = \mathbb{Z}/n\mathbb{Z}$$

so this homomorphism is surjective.

Example. Consider the "evaluate at" map

$$ev_\alpha : \mathbb{Q}[x] \to \mathbb{Q} \qquad ev_\alpha(f) = f(\alpha)$$

This is a homomorphism due to properties of polynomials

$$(f + g)(\alpha) = f(\alpha) + g(\alpha) \qquad fg(\alpha) = f(\alpha)g(\alpha)$$

We also have

$$\ker ev_\alpha = \{(x-\alpha)f \mid f \in \mathbb{Q}[x]\}$$

The kernel is the set of polynomials with α as a root. This homomorphism is also surjective since we can construct a polynomial that evaluates to any rational number, namely we can take the set of constant polynomials.

Example. If ρ is a complex representation of S_3, we can construct a ring homomorphism

$$\phi : \mathbb{C}[S_3] \to M_n(\mathbb{C}) \qquad \phi\left(\sum_{\sigma\in S_3} a_\sigma \sigma\right) = \sum_{\sigma\in S_3} a_\sigma \rho(\sigma)$$

When discussing groups, we noted that the kernel of a homomorphism was a normal subgroup and thus we can take a quotient using it. The analogous structure for rings is the ideal.

Definition. *Let R be a ring, a nonempty subset $I \subseteq R$ is an ideal if*

1. *$(I,+) \leq (R,+)$, it is an additive subgroup*
2. *For any $a \in I$, $ra \in I$ for all $r \in R$*

This defines a left subgroup because it absorbs multiplication from the left, we define right ideals in a similar way. If an ideal is both left and right, it is called a two-sided ideal or simply just ideal.

Naturally for commutative rings, left ideals, right ideals, and ideals are all the same. Given this we can easily posit a set of conditions which must be satisfied for a set to be an ideal.

Proposition 5.7. Let R be a ring, a subset $I \subseteq R$ is an ideal if

1. $I \neq \varnothing$
2. $\forall a, b \in I; a+b \in I$
3. For $a \in I, r \in R; ar, ra \in I$

Example. $n\mathbb{Z}$ is an ideal of $\mathbb{Z}$

Example. Let R be a commutative ring and $\alpha \in R$, then the set

$$\{(x-\alpha)g \mid g \in R[x]\}$$

is an ideal of $R[x]$.

Example. Let $R = M_2(\mathbb{R})$ be the ring of 2×2 matrices and consider

$$I = \left\{ \begin{pmatrix} 0 & a \\ 0 & b \end{pmatrix} \middle| \, a, b \in \mathbb{R} \right\}$$

This is a left ideal of $M_2(\mathbb{R})$

Since ideals are analogous to normal subgroups we can now define an analogous version of simple groups.

Definition. *A ring is simple if it has no nontrivial two-sided ideals.*

Proposition 5.8. A simple commutative ring is a field

Proof. Let R be a commutative ring, then the fact that it is simple means it can have no nontrivial ideals. Suppose $x \in R$ and consider the ideal (x) which must be the entire ring so $1 \in (x)$. This means that there is some $r \in R$ such that $xr = 1$ which means x is a unit, thus R is a field. □

We can now use the idea of ideals to define quotient rings.

Definition. *Let R be a ring and I an ideal of R, the quotient ring is the set*

$$R/I = \{r + I \mid r \in R\}$$

We can think of this a set of left cosets of the additive group $(I, +)$.

Proposition 5.9. R/I is a ring under the operations

$$(r + I) + (s + I) = (r + s) + I$$
$$(r + I)(s + I) = rs + I$$

There a surjective map $\pi : R \to R/I$ given by $\pi(r) = r + I$. This map is called the projection map.

Now we can formulate the isomorphism theorems for rings.

Proposition 5.10. Let $\varphi : R \to S$ be a ring homomorphism, then

1. $ker\,\varphi$ is an ideal of R.
2. $im\,\varphi$ is a subring of S

Proof. $\varphi(0) = 0$ so it is nonemply, let $a, b \in ker\,\varphi$ and $r \in R$, then

$$\varphi(a + b) = \varphi(a) + \varphi(b) = 0$$
$$\varphi(ar) = \varphi(a)\varphi(r) = 0$$
$$\varphi(ra) = \varphi(r)\varphi(a) = 0$$

So $a + b, ar, ra \in \ker\varphi$ and thus it is an ideal, in fact it is a two-sided ideal.

The image is nonempty because R is nonempty, let $a, b \in \text{im } \varphi$ and $x, y \in R$ such that

$$\varphi(x) = a \qquad \varphi(y) = b$$

then

$$\varphi(x - y) = \varphi(x) - \varphi(y) = a - b$$
$$\varphi(xy) = \varphi(x)\varphi(y) = ab$$

So $a - b, ab \in \text{im } \varphi$ and it is a subring. □

Theorem 5.11 (The Isomorphism Theorems)**.** Here we will list all four isomorphism theorems

1. If $\varphi : R \to S$ is a homomorphism, then
$$R/ker\,\varphi \cong \text{im } \varphi$$
2. Let A be a subring and I an ideal of R, then
$$A + I = \{a + b \mid a \in A, b \in I\} \subseteq R$$
is a subring, $A \cap I$ is an ideal, and
$$(A + I)/I \cong A/(A \cap I)$$
3. Let I, J be ideals of R with $I \subseteq J$, then J/I is an ideal of R and
$$(R/I)/(J/I) \cong R/J$$
4. Let R be a ring and I an ideal, then there exists a bijection
$$\{\text{ideals of } R/I\} \Longleftrightarrow \{\text{ideals of } R \text{ containing } I\}$$

When discussing ideals, we will often consider ideals generated by one or more elements similar to how we can construct subgroups using generators. If $a_1, \ldots, a_n \in R$, then the ideal generated by these elements is denoted
$$(a_1, \ldots, a_n) = \left\{\sum r_i a_i \mid r_i \in R\right\}$$

Example. Using the evaluation map ev_α we can construct
$$R/(x - \alpha) \cong R$$
using the first isomorphism theorem.

Example. We discussed the Gauss integers $\mathbb{Z}[i]$ previously as a subring of $\mathbb{C}$, we can also view this ring as a quotient ring.
$$\mathbb{Z}[x]/(x^2 + 1) \cong \mathbb{Z}[i]$$
Modding by the ideal generated by x^2+1 is equivalent to stipulating $x^2+1 = 0$ which defines the imaginary numbers. The elements are thus of the form $a + bx$ where $x^2 + 1 = 0$, which is the Gauss integers.

5.4 More on Ideals

It is possible to form new ideals from known ideals. If I, J are ideals, then
$$I + J = \{a + b \mid a \in I, b \in J\}$$
$$IJ = \left\{\sum a_i b_i \mid a_i \in I, b_i \in J\right\}$$
are both ideals. The intersection $I \cap J$ is also an ideal, furthermore $IJ \subseteq I \cap J$. We can use this fact to prove the following useful theorem. It is usually stated in number theoretic terms using modulo, but the more general case is presented here.

Theorem 5.12 (Chinese Remainder Theorem)**.** Let R be a commutative ring with $1 \neq 0$ and I, J be ideals of R such that $I + J = R$, then

1. $IJ = I \cap J$
2. $R/IJ = R/(I \cap J) \cong R/I \times R/J$

Proof. 1) Since $I + J = R$, this implies that there exists elements $a \in I$, $b \in J$ such that $a + b = 1$. Consider some $x \in I \cap J$, we can write this as

$$x = x \cdot 1 = x \cdot (a + b) = x \cdot a + x \cdot b \in I + J$$

Thus $I \cap J \subseteq IJ$, the other inclusion is trivial due to properties of ideals, so we can conclude that $IJ = I \cap J$.

2) Consider a map $\phi : R \to R/I \times R/J$ given by

$$\phi(x) = (x + I, x + J)$$

Suppose $x \in \ker \phi$ so that

$$\phi(x) = (x + I, x + J) = (I, J)$$

This implies that $x \in I, J$ so $x \in I \cap J$. For the other direction suppose $x \in I \cap J$, then

$$(I, J) = (x + I, x + J) = \phi(x)$$

and so $x \in \ker \phi$. Thus $\ker \phi = I \cap J$. Consider some $(x+I, y+J) \in R/I \times R/J$ and suppose $a \in I, b \in J$ such that $a + b = 1$. Then

$$\begin{aligned} \phi(bx + ay) &= (bx + ay + I, bx + ay + J) \\ &= (bx + I, ay + J) \\ &= ((1 - a)x + I, (1 - b)J) \\ &= (x + I, y + J) \end{aligned}$$

Thus ϕ is surjective and we can conclude by the first isomorphism theorem that

$$R/(I \cap J) \cong R/I \times R/J$$

□

Example. Consider the ring $\mathbb{R}[x]$ and the ideal $I = (x^2 - 1)$. Note that we can factor

$$x^2 - 1 = (x + 1)(x - 1)$$

Furthermore we can write

$$1 = \frac{1}{2}(x + 1) - \frac{1}{2}(x - 1) \in (x - 1) + (x + 1)$$

Thus $(x + 1) + (x - 1) = \mathbb{R}[x]$, so by the Chinese Remainder Theorem

$$\mathbb{R}[x]/(x^2 - 1) \cong \mathbb{R}[x]/(x + 1) \times \mathbb{R}[x]/(x - 1) \cong \mathbb{R} \times \mathbb{R}$$

Recall that for a set of elements $\{a_1, \ldots, a_n\}$ we denote the ideal generated by them with $(a_1, \ldots, a_n)$. This notion can be extended to an arbitrary set A and the ideal generated by A, denoted (A), is the smallest ideal containing all of A.

Definition. *An ideal generated by a single element $I = (a)$ is called principle.*

Proposition 5.13. Let R be a ring with $1 \neq 0$ and I an ideal

1. $I = R$ if and only if I has a unit
2. If R is commutative, then it is a field if and only if the only ideals of R are the trivial ones, $(0), R$
3. Let F be a field and $\phi : F \to R$ be a nonzero ring homomorphism, then ϕ is injective.

Proof. 1) If $I = R$, then $1 \in R$ which is a unit. Conversely let $r \in R$ and $u \in I$ be a unit, then

$$r \cdot 1 = r(uu^{-1}) = (ru^{-1})u \in I$$

Thus $I = R$.

2) Let R be a field and I an ideal. Suppose $I \neq (0)$ which means there is some nonzero element in I. Since R is a field it must be a unit and so by (1), $I = R$. Conversely suppose the only two ideals are $I = (0), R$. Then for any nonzero element u, the ideal generated is $(u) = R$ and so $1 \in (u)$. This implies that there is some $v \in R$ such that $uv = 1$ and so u is a unit. This is true for all nonzero elements, so R is a field.

3) Consider some $\phi : F \to R$. The kernel is an ideal so it is either (0) or R. If it is R, then the map is zero and if it is (0) then it is injective. □

Now we will discuss some special types of ideals.

Definition. *An ideal M of a ring R is maximal if $M \neq R$ and for any other ideal N, $N \subset M$. In other words the only ideals containing M are M and R.*

Proposition 5.14. Let R be a ring with $1 \neq 0$, then R contains a maximal ideal. In particular, every proper ideal is contained within a maximal ideal.

Proof. Let Σ be the set if of all proper ideals in R, which is nonempty and can be ordered by inclusion. Suppose we have a chain

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots \subseteq I_n$$

then let $I = \bigcup I_\alpha$ is an ideal which contains all the ideals in the chain. Note that since $1 \notin I_\alpha$ for all α then $1 \notin I$, this this forms an upper bound for the chain. By Zorn's lemma there exists an maximal element of Σ which is the maximal ideal.

If we have some ideal J, then we can find the maximal ideal of R/J which is a maximal ideal of R which contains J using the fourth isomorphism theorem. □

By modifying the proof slightly and letting Σ be the set of proper ideals which contain a non unit, we can show that

Corollary 5.15. Every non-unit is contained in a maximal ideal

There is an easy way to characterise maximal ideals by forming quotients

Proposition 5.16. Let R be commutative and I an ideal. I is maximal if and only if R/I is a field

Proof. R/I is a field if and only if (0) and R/I are its only ideals. This occurs if and only if the only ideals containing I are I and R, which would imply that I is maximal. □

Example. The ideal $(n) \subseteq \mathbb{Z}$ is maximal if $|n|$ is prime. This implies that $\mathbb{Z}/n\mathbb{Z}$ is a field if and only if n is prime.

Example. The ideal $(2, x) \subseteq \mathbb{Z}[x]$ is maximal and

$$\mathbb{Z}[x]/(2, x) \cong \mathbb{F}_2$$

where $\mathbb{F}_2 = \{0, 1\}$. To see this consider the map

$$\phi : \mathbb{Z}[x] \to \mathbb{F}_2 \qquad \phi(f) = f(0) \mod 2$$

Lets compute the kernel of this homomorphism. If $f \in \ker \phi$, then

$$f = \sum a_i x_i \qquad \phi(f) = \overline{f(0)} = \overline{a_0} = 0 \to a = 2b$$

for some integer b. Thus we can rewrite

$$f(x) = 2b + \sum_{i=1} a_i x_i = 2b + x \sum_{i=0} a_i x^i = 2g(x) + xh(x) \in (2, x)$$

This implies that $\ker \phi \subseteq (2, x)$. Now for the other direction suppose $f \in (2, x)$, then

$$\phi(f) = \overline{2f(0)} + \overline{0g(0)} = \overline{2} = 0$$

Thus $(2, x) \subseteq \ker \phi \to \ker \phi = (2, x)$. This map is quite obiously surjective because

$$\phi(0) = 0 \qquad \phi(1) = 1$$

Thus by the first isomorphism theorem

$$\mathbb{Z}[2]/(2, x) \cong \mathbb{F}_2$$

and so the ideal $(2, x)$ is maximal.

Example. Consider the ring of one way infinite series

$$\mathbb{C}[[x]] = \left\{ \sum_{i=0}^{\infty} a_i x^i \middle| a_i \in \mathbb{C} \right\}$$

Addition is done traditionally and multiplication is done by

$$\left(\sum_{i=0}^{\infty} a_i x^i \right) \left(\sum_{i=0}^{\infty} b_i x^i \right) = \sum_{i=0}^{\infty} c_i x^i$$

where $c_n = \sum_{i+j=n} a_i b_j$. The only invertible elements of $\mathbb{C}[[x]]$ are the polynomials with a nonzero constant term. This ring has a unique maximal ideal consisting of all the non-invertible elements. If you mod out all such elements, you are left with only the constant polynomials $\mathbb{C}$ which is obviously a field.

In the previous example we have a ring with a unique maximal ideal, these rings are called local rings. A ring with only a finite number of maximal ideals is called semi-local. If a ring R has a unique maximal ideal M, we call the resulting field $k = R/M$ its residue field.

Example. A useful example of a local ring involves units. Let R be a ring and M an ideal such that every element $x \in R \setminus M$ is a unit, then R is local with unique maximal ideal M. This is because every proper ideal must consist only of non-units, so they will be contained in M.

Definition. *Let R be a commutative ring with $1 \neq 0$. An ideal I is prime if $I \neq R$ and if $ab \in I$ then either $a \in I$ or $b \in I$.*

There is also an easy way to see if an ideal is prime.

Proposition 5.17. An ideal $I \subseteq R$ is prime if and only if R/I is an integral domain.

Proof. Suppose $I \neq R$, then if I is prime, $ab \in I$ would imply $a \in I$ or $b \in I$. This means that in the quotient R/I, $\overline{ab} = 0$ implies $\overline{a} = 0$ or $\overline{b} = 0$. But this is just the definition of zero divisors and this property means that zero divisors cannot exist so R/I is an integral domain. □

Corollary 5.18. If R is a commutative ring and I is a maximal ideal, then I is also prime.

Proof. If I is maximal, then R/I is a field which is also an integral domain. □

Example. The ideal $(n) \subseteq \mathbb{Z}$ is prime if $n = 0$ or $|n|$ is prime. This is because $\mathbb{Z}/(n)$ is a field if $|n|$ is prime. The ideal (0) is prime but not maximal.

From this example we see that the notion of a prime ideal is in some ways a generalization of the familiar prime numbers.

Example. The ideal $(x) \in \mathbb{Z}[x]$ is prime but not maximal since

$$\mathbb{Z}[x]/(x) \cong \mathbb{Z}$$

but $(x) \subset (2, x)$

Example. Consider the polynomial ring $\mathbb{R}[x, y]$ and consider the ideal $I = (y - x^2)$. This ideal is prime, suppose we define the map

$$\phi : \mathbb{R}[x, y] \to \mathbb{R}[x] \qquad \phi(f(x, y)) = f(x, x^2)$$

We see that $\phi(y - x^2) = 0$ and so the kernel is generated by this. This map is clearly surjective because $\mathbb{R}[x] \subset \mathbb{R}[x, y]$ and so

$$\mathbb{R}[x, y]/(y - x^2) \cong \mathbb{R}[x]$$

The polynomial ring is an integral domain so I is prime. As another example, the ideal (xy) is not prime because $x, y \notin (xy)$.

Prime ideals can be used to give a concept of dimensionality to rings.

Definition. *The length of the largest (proper) chain of primes ideals of a ring*

$$P_0 \subset P_1 \subset \cdots \subset P_n$$

is called the Krull dimension.

Example. Here are Krull dimensions for some common rings

- $\dim \mathbb{Z} = 1$
- $\dim k = 0$ for algebraically closed k
- $\dim k[x] = 0$
- $\dim k[x_1, \ldots, x_n] = n$

For a commutative ring, the set of all prime ideals is called its spectrum, denoted

$$\text{Spec } A = \{P \mid P \subseteq A \text{ prime}\}$$

this can be given a topology, called the Zariski topology, and is used in algebraic geometry.

Definition. *An element $x \in A$ is nilpotent if $x^n = 0$ for some $n \geq 1$. The set of nilpotent elements forms an ideal*

$$\mathfrak{N} = \{x \in A, x^n = 0 \text{ for some } n \geq 1\}$$

called the nilradical.

The fact that this is an ideal can be seen by using the binomial expansion theorem. Note that a nilpotent element will also be a zero divisor but not vice versa, thus a nilpotent element will be contained in every prime ideal. Using this fact we can construct the nilradical in another way.

$$\mathfrak{N} = \bigcap_{P \in \text{Spec } A} P$$

Sometimes it is not desirable to have nilpotent elements, we call such a ring without nilpotent elements reduced. In other words, a reduced ring will have a trivial nilradical $\mathfrak{N} = (0)$. The advantage of defining the nilradical is that by taking the quotient of a ring with its nilradical, we can create a reduced ring, which is called its reduction.

Proposition 5.19. The reduction $A/\mathfrak{N}$ is reduced.

Proof. Let $\overline{x} \in A/\mathfrak{N}$ be represented by $x \in A$, then clearly $\overline{x}^n$ would be represented by x^n. It then follows that

$$\overline{x}^n = 0 \to x^n \in \mathfrak{N} \to (x^n)^k = 0 \text{ for some } k > 0 \ \to x \in \mathfrak{N} \to \overline{x} = 0$$

thus there are no nontrivial nilpotent elements. □

If we were to take the intersection of all maximal ideals rather than all prime ideals, we again get another ideal: the Jacobson radical $\mathfrak{R}$. This is also characterized another way

Proposition 5.20. An element $x \in \mathfrak{R}$ if and only if $1 - xy$ is a unit for all $y \in A$.

Proof. Suppose for the purpose of contradiction that $1 - xy$ is not a unit, this means that it must belong in some maximal ideal M. But $x \in \mathfrak{R}$ implies that $x \in M$ and so $xy \in M \to 1 \in M$, a clear contradiction.

Conversely suppose $x \notin \mathfrak{R}$, that is there exists some maximal ideal M such that $x \notin M$. Then M and x generate the unit ideal (1) so that $u + xy = 1$ for some $u \in M$ and $y \in A$. Thus $1 - xy \in M$ and is therefore not a unit. □

We will finish our lengthy discussion on ideals by discussing some more operations. First we have a useful theorem concerning intersections and unions of ideals.

Proposition 5.21. i) Let $\mathfrak{p}_1, \cdots, \mathfrak{p}_n$[2] be prime ideals and $\mathfrak{a}$ an ideal contained within their union. Then $\mathfrak{a} \subseteq \mathfrak{p}_i$ for some i.

$$\mathfrak{a} \subseteq \bigcup_{i=1}^{n} \mathfrak{p}_i \quad \to \quad \mathfrak{a} \subseteq \mathfrak{p}_i \text{ for some } i$$

ii) Let $\mathfrak{a}_1, \cdots, \mathfrak{a}_n$ be ideals and $\mathfrak{p}$ a prime ideal containing their intersection, then $\mathfrak{p} \supseteq \mathfrak{a}_i$ for some i. If $\mathfrak{p}$ is equal to the intersection then this containment becomes equality.

$$\mathfrak{p} \supseteq \bigcap_{i=1}^{n} \mathfrak{a}_i \quad \to \quad \mathfrak{p} \supseteq \mathfrak{a}_i \text{ for some } i$$

Proof. Omitted [3] □

The first new operation we introduce is the ideal quotient, in the field of algebraic geometry this corresponds to the operation of set difference (set minus).

Definition. *Let $\mathfrak{a}, \mathfrak{b}$ be ideals in a ring A, their ideal quotient is*

$$(\mathfrak{a} : \mathfrak{b}) = \{x \in A \mid x\mathfrak{b} \subseteq \mathfrak{a}\}$$

[2] This is another way to denote ideals which I will sometimes use, it is common in commutative algebra and algebraic geometry

[3] Lazy, see Atiyah-MacDonald for proof

This is also an ideal, in particular the quotient $(0 : \mathfrak{b})$ is called the annihilator (of $\mathfrak{b}$) and denoted $Ann(\mathfrak{b})$. In the case that we deal with a principle ideal we may write

$$(x : \mathfrak{b}) = ((x) : \mathfrak{b})$$

Using this notation, the set of zero divisors is the union of $Ann(x)$ for all nonzero x.

Proposition 5.22. The ideal quotient has the following properties

1. $\mathfrak{a} \subseteq (\mathfrak{a} : \mathfrak{b})$
2. $(\mathfrak{a} : \mathfrak{b})\mathfrak{b} \subseteq \mathfrak{a}$
3. $((\mathfrak{a} : \mathfrak{b}) : \mathfrak{c}) = (\mathfrak{a} : \mathfrak{b}\mathfrak{c}) = ((\mathfrak{a} : \mathfrak{c}) : \mathfrak{b})$
4. $(\cap_i \mathfrak{a}_i : \mathfrak{b}) = \cap_i (\mathfrak{a}_i : \mathfrak{b})$
5. $(\mathfrak{a} : \sum_i \mathfrak{b}_i) = \cap_i (\mathfrak{a} : \mathfrak{b}_i)$

Proof. Omitted [3] □

The radical of an ideal is another way to form new ideals.

Definition. *The radical of an ideal $\mathfrak{a} \subseteq A$ is*

$$r(\mathfrak{a}) = \sqrt{\mathfrak{a}} = \{x \in A \mid x^n \in \mathfrak{a} \text{ for some } n > 0\}$$

We can see that this is indeed an ideal by noting that it is the preimage of the nilradical (of $A/\mathfrak{a}$) under the standard homomorphism.

Proposition 5.23. The ideal radical has the following properties

1. $r(\mathfrak{a}) \subseteq \mathfrak{a}$
2. $r(r(\mathfrak{a})) = r(\mathfrak{a})$
3. $r(\mathfrak{a}\mathfrak{b}) = r(\mathfrak{a} \cap \mathfrak{b}) = r(\mathfrak{a}) \cap r(\mathfrak{b})$
4. $r(\mathfrak{a}) = 1$ if and only if $\mathfrak{a} = (1)$
5. $r(\mathfrak{a} + \mathfrak{b}) = r(r(\mathfrak{a}) + r(\mathfrak{b}))$
6. If $\mathfrak{p}$ is prime, then $r(\mathfrak{p}^n) = \mathfrak{p}$ for all $n > 0$

Proof. Omitted [3] □

The fact that the radical of an ideal is the preimage of the nilradical allows us to define the radical in terms of prime ideals similar to how the nilradical is defined.

[3] There's too many and I'm lazy, the actual proof is an exercise in Atiyah-MacDonald

Proposition 5.24. The radical of an ideal $\mathfrak{a}$ is the intersection of all prime ideals which contain $\mathfrak{a}$.

Proof. Consider the quotient ring $A/\mathfrak{a}$, the radical $r(\mathfrak{a})$ is the same as the nilradical of $A/\mathfrak{a}$ and so can be identified as the intersection of all prime ideals of $A/\mathfrak{a}$. This is the same as the intersection of prime ideals which contain $\mathfrak{a}$. □

As a final point of discussion, ideals can be carried over a homomorphism. Let $f : A \to B$ be a homomorphism of rings and $\mathfrak{a}$ an iddeal in A. In general the set $f(\mathfrak{a})$ is not an ideal however we can make it one by considering all the elements it generates. The extension of an ideal is defined as

$$\mathfrak{a}^e = Bf(\mathfrak{a}) = \left\{ \sum y_i f(x_i) \mid x_i \in \mathfrak{a}, y_i \in B \right\}$$

The other direction is way easier, if $\mathfrak{b}$ is an ideal in B then $f^{-1}(\mathfrak{b})$ will always be an ideal of A. We will denote this ideal, called the contraction, with $\mathfrak{b}^c$. Note that if $\mathfrak{b}$ is prime then so is its contraction, but if $\mathfrak{a}$ is prime the extension may not be.

5.5 Special Rings

There are some special rings with more structure than generic rings which we would like to discuss in more detail here, they will assist us in later discussions. All rings are assumed to be commutative in this section.

Definition. *For an integral domain R, a norm is a function $N : R \to \mathbb{Z}^+ \cup \{0\}$ with $N(0) = 0$. If $N(r) > 0$ for all $r \neq 0$, then N is a positive norm.*

What we've done here is essentially defined a measure of size for some arbitrary ring element. The way we choose to define our norm is up to us, indeed there can be multiple norms on the same integral domain.

Definition. *An integral domain R is a Euclidean Domain if there is a norm N on R such that for any $a, b \in R$ with $b \neq 0$, there exists $q, r \in R$ such that*

$$a = qb + r \qquad r = 0 \text{ or } N(r) < N(b)$$

q is called the quotient and r the remainder of division.

Thus a Euclidean Domain is a ring in which we can use division in the familiar sense. This means that we can use the Euclidean algorithm discussed in Chapter 1.

Example. Fields are trivially Euclidean Domains because any norm will satisfy the condition. For any a, b with $b \neq 0$, we have $a = qb$ where $q = ab^{-1}$

Example. The integers have a norm given by $N(x) = |x|$, allowing us to divide integers which is great because we've been doing this since elementary school.

Proposition 5.25. Every ideal in a Euclidean Domain is principle.

Proof. Suppose I is a nonzero ideal and let $d \in I$ be the element with minimum norm. It's obvious that $(d) \subseteq I$ because $d \in I$. For the other direction let $a \in I$. Since this is a Euclidean Domain, we can write $a = qd + r$ where $r = 0$ or $N(r) < N(d)$. But we can write $r = a - qd \in I$ and since d has minimum norm in I it must be the case that $r = 0$. Thus $a = qd \in (d)$ and so $I = (d)$ and is principle. □

Such a ring in which every ideal is principle is also given a special name.

Definition. *A ring in which every ideal is principle is known as a Principle Ideal Domain (PID).*

Example. The integers $\mathbb{Z}$ are a PID, every ideal is of the form $n\mathbb{Z} = (n)$.

PIDs will be important in later discussions about modules so lets first prove some useful results.

Definition. *Let R be a commutative ring and $a, b \in R$ with $b \neq 0$.*

1. *a is a multiple of b if there exists some $q \in R$ such that $a = qb$. In this case we say that b is a divisor of a.*

2. *The greatest common divisor of a, b is some nonzero d dividing both a and b and if any other d' has this property then d' divides d.*

These definitions should be familiar as they are just ring theoretic versions of their number theoretic counterparts. The notion of a greatest common divisor is useful when discussing ideals.

Proposition 5.26. If a, b are nonzero elements of some ring R which generate a principle ideal $(a, b) = (d)$, then d is the greatest common divisor of a and b.

Proof. If we translate the definition of GCD into the language of ideals, we have that d is the GCD of a and b if the ideal $I = (a, b)$ is contained within (d) and if (d') contains I, then $(d) \subseteq (d')$. In other words the ideal generated by the GCD is the smallest ideal containing both a and b. □

For some general ring the greatest common divisor may not be unique.

Proposition 5.27. Let R be an integral domain and suppose two elements $d, d' \in R$ both generate the same principle ideal. Then $d' = ud$ for some unit $u \in R$. In particular if both d, d' are greatest common divisors then they are multiples of each other by some unit.

Proof. Suppose d, d' are nonzero. Since $d \in (d')$ we can write $d = xd'$ for some $x \in R$. By a similar argument we can also write $d' = yd$ for some $y \in R$. This implies that

$$d = xd' = xyd \to d(xy - 1) = 0 \to xy = 1$$

Thus x, y are units. □

These two propositions combine to form a very useful property of PIDs. Before stating the theorem it is worth noting that while every Euclidean Domain is a PID not every PID is a Euclidean Domain.

Theorem 5.28. Let R be a PID and a, b be nonzero elements. Let d be their GCD, then d can be written as an R-linear combination, that is there exists $x, y \in R$ such that

$$d = ax + by$$

Proof. This is just an application of the previous propositions and the Euclidean algorithm □

Finally, one more useful property of PIDs.

Proposition 5.29. Every nonzero prime ideal in a PID is maximal.

Proof. Let (p) be a nonzero prime ideal and suppose $I = (m)$ contains (p). This means that $p \in (m)$ and so $p = rm$ for some $r \in R$. Since (p) is prime, this implies that either $r \in (p)$ or $m \in (p)$. If $m \in (p)$, then the two ideals are the same. If $r \in (p)$, then

$$r = ps \to p = rm = psm \to sm = 1$$

So m is a unit and $I = R$. Thus (p) is maximal. □

Corollary 5.30. If R is a commutative ring such that $R[x]$ is a PID, then R is a field

Proof. Since R is a subring of $R[x]$ it must also be an integral domain. We can express R is terms of a quotient

$$R \cong R[x]/(x)$$

which would imply that the ideal (x) is prime which means it is also maximal by the previous proposition. This means that R is a field. □

So we've seen that Euclidean Domains are also Principle Ideal Domains. There is one more ring with even more structure, namely that elements can be factored. But first we need to define what that even means.

Definition. *Let R be an integral domain*

1. *Suppose $r \in R$ is nonzero and not a unit, then r is irreducible if whenever $r = ab$, either a or b is a unit. Otherwise r is reducible.*
2. *A nonzero $p \in R$ is prime if (p) is a prime ideal.*
3. *Two elements differing by a unit ($a = ub$, u is a unit) are associate.*

These definitions are analogous versions of familiar concepts lifted over to rings.

Proposition 5.31. In an integral domain a prime element is always irreducible.

Proof. Suppose (p) is a prime ideal and $p = ab \in (p)$. Then either a or b is in (p), let's say it's a, and so $a = pr$ for some r. Thus

$$p = ab = prb \to rb = 1$$

so b is a unit which means p is irreducible. □

Proposition 5.32. In a PID a nonzero element is prime if and only if it is irreducible

Proof. A PID is an integral domain so one direction is already shown. For the other direction suppose p is irreducible and let I be an ideal containing (p). In a PID, every ideal is principle so $I = (r)$ for some r in the PID. Since $p \in I$ this means that $p = rm$ for some r. But p is irreducible so either r or m is a unit. If r is a unit then $m = r^{-1}p$ and $(p) = (m)$. If m is a unit then $(m) = (1)$ which means that the ideal (p) is maximal. Maximal ideals in a PID are also prime thus completing the proof. □

Definition. *A Unique Factorization Domain (UFD) is an integral domain R such that for every nonzero $r \in R$ which is not a unit...*

1. *r can be factored in terms of irreducibles $p_i \in R$*

$$r = p_1 p_2 \cdots p_n$$

2. *This decomposition is unique up to associates.*

Example. Any field is trivially a UFD since all nonzero elements are units.

It can be shown that PIDs are UFDs which mean that properties of both PIDs and Euclidean Domains hold in a UFD.

Theorem 5.33. Every Principle Ideal Domain is a Universal Factorization Domain.

Proof. Let R be a PID and $r \in R$ a nonzero element which is not a unit. We want to show that r can be expressed as a product of irreducible elements and that this decomposition is unique up to units. Suppose r is reducible because otherwise we are done. By definition we can write $r = r_1 r_2$ where neither are units. If both elements are irreducible then we are done, otherwise we can write one of them (say r_1 for instance) as $r_1 = r_{11} r_{12}$ and so forth. We must show that this process terminates so that every factor of r is irreducible.

Suppose this process doesn't terminate. From the factorization $r = r_1 r_2$ we get the inclusion of ideals

$$(r) \subset (r_1) \subset R$$

Note that inclusions are proper because neither r_1 nor r_2 are units. If the process doesn't terminate then we get the infinite ideal chain

$$(r) \subset (r_1) \subset (r_{11}) \subset \cdots \subset R$$

Let's denote the infinite ideal chain as

$$I_1 \subset I_2 \subset I_3 \subset \cdots \subset R$$

Let $I = \bigcup I_i$ which is an ideal and, since R is a PID, suppose $I = (a)$. But I is the union of infinitely many ideals so a must be in one of them. Let n be such that $a \in I_n$ and note that

$$I_n \subseteq I = (a) \subseteq I_n$$

Thus $I_n = I$ and the chain becomes stationary at n. This proves that every nonzero element of R which is not a unit has some factorization in terms of irreducible elements.

Now we must prove that this is unique up to multiplication by a unit. We will do this through induction, let n be the number of irreducible factors in a factorization of some element r. If $n = 0$ then r is a unit and we are done. For $n = 1$, r is irreducible. If there is some other factorization $r = qc$ where q is irreducible, then c is a unit and q divides a unit, making it a unit which is a contradiction.

Suppose now that $n \geq 1$ and we have the factorizations

$$r = p_1 p_2 \cdots p_n = q_1 q_2 \cdots q_m \quad m \geq n$$

where p_i and q_j are not necessarily distinct. This implies that p_1 will divide the right product, without loss of generality suppose it divides q_1. But these are irreducible, so $q_1 = p_1 u$ for some unit u and so they are associate. Since we are in an integral domain, we can cancel factors of p_1 to get

$$p_2 p_3 \cdots p_n = u q_2 q_3 \cdots q_m = q_2' q_3 \cdots q_m \quad m \geq n$$

where $q_2' = u q_2$ is irreducible. By induction on n, we conclude that factors on the left and right sides match up to associates. Since p_1 and q_1 are already associate, this completes the inductive step and thus the theorem. □

6 Introduction to Modules

6.1 Definition and Properties

A module is an algebraic structure which generalizes the notion of a vector space. In a sense it is a structure with an associated "ring" action on an underlying group. For the next few sections in chapters we will asume that all rings have a multiplicative identity $1 \neq 0$.

Definition. *Let R be a ring, a left R-module is an abelian group M with an action*

$$\cdot : R \times M \to M$$

such that for $r, s \in R$, $m, n \in M$

1. $r \cdot (m + n) = rm + rn$

2. $(r+s)\cdot m = rm + sm$

3. $(rs)\cdot m = r\cdot(sm)$

4. $1\cdot m = m$

A right module has the same definition but with an action

$$\cdot : M \times R \to M$$

A bimodule is compatibly both a left and right module

$$r\cdot s\cdot m = (rm)\cdot s = r\cdot(ms)$$

Sometimes the following convention is used to make discussions easier

- M_B is a right B-module
- ${}_AM$ is a left A-module
- ${}_AM_B$ is a A-B-bimodule

Note that M_B and ${}_{\mathbb{Z}}M_B$ are the same since every module can be trivially extended to a bimodule through $\mathbb{Z}$. By convention modules are left acting unless otherwise specified. If R is a field, then an R-module is just a vector space over R.

Definition. *Let M be an R-module. A subgroup $N \leq M$ is a submodule if $rn \in N$ for all $r \in R, n \in N$.*

For every module M, we have the trivial submodules M and $\{0\}$.

Example. A ring is bimodule over itself by multiplication, it can also be thought of as a left/right module by only considering left/right multiplication. The left/right submodules are its left/right ideals and a sub-R-bimodule is a two-sided ideal.

Example. Let $\varphi : R \to S$ be a ring homomorphism and suppose M is a S-module. Then M is also an R-module with the action

$$R \times M \to M \qquad (r, m) \mapsto f(r)m$$

Example. Consider the ring of integers $\mathbb{Z}$ and any abelian group A. We can define the action

$$na = \begin{cases} a + a + \cdots + a \ (n \text{ times}) & n > 0 \\ 0 & n = 0 \\ -a - a - \cdots - a \ (-n \text{ times}) & n < 0 \end{cases}$$

This action makes A into a $\mathbb{Z}$-module. Since modules are by definition abelian groups this implies that a $\mathbb{Z}$-module is just an abelian group. The submodules are just subgroups of A.

As usual there is a simple criterion for submodules.

Proposition 6.1. Let R be a ring and M an R-module. A subset $N \subseteq M$ is a submodule if and only if

1. N is nonempty
2. $x + ry \in N$ for all $r \in R$, $x, y \in N$

Proof. If $r = -1$, then this is just the subgroup test. For $x = 0$, this ensures that the module is closed under the ring action, thus it is a submodule. □

Example. Consider a complex polynomial ring $\mathbb{C}[x]$ and suppose we constructed a $\mathbb{C}[x]$-module. Let V be a $\mathbb{C}$-module, which is a vector space over $\mathbb{C}$, and let $T : V \to V$ be a linear transform. We can define the ring action as

$$f(x)v = \left(\sum a_i x^i\right) v = \sum a_i T^i(v)$$

The choice of T is completely arbitrary as long as it is linear, so a $\mathbb{C}[x]$-module is just a complex vector space equipped with a linear transform. Such a transform from $V \to V$ is called an endomorphism. A submodule is a subspace $W \subseteq V$ which is invariant under T, that is $T(W) \subseteq W$.

Example. Consider a group ring $\mathbb{C}[G]$ where G is some finite group. A $\mathbb{C}[G]$-module constains a $\mathbb{C}$-module structure using the action

$$zm = (z1_G)m$$

Let $g, h \in G$ be group elements, they act on the complex vector space V as

$$(gh)v = g(hv) \qquad g(v_1 + v_2) = gv_1 + gv_2$$

In other words g is just a linear transformation $V \to V$. This gives a homomorphism

$$\rho : G \to GL(V)$$

which is just a complex representation. Thus a $\mathbb{C}[G]$-module is a complex representation of G, in particular it is the complex vector space V equipped with a homomorphism ρ. Naturally a submodule is just a subrepresentation.

Definition. *Let R, k be rings with k commutative. A k-algebra structure on R is a map*

$$k \to Z(R) = \{x \in R \mid xy = yx \ \forall y \in R\}$$

When we say "a k-algebra R," we are really saying "let R be a ring with a k-algebra structure." Thus the elements of k act on and commute with elements of R.

Example. Any ring with identity is a $\mathbb{Z}$-algebra because we can define a homomorphism by the fact that 1 is sent to the ring identity.

Example. Suppose $R = F$ is a field. Ring homomorphisms from fields are injective so $F \cong f(F)$, in other words a F-algebra A would contain an isomorphic copy of F as a subring. So an algebra over a field is just any ring containing that field, this is sometimes called a K-algebra.

6.2 Homomorphisms and Quotient Modules

Continuing with the usual path of discussion for algebraic structures we now define the notion of maps between structures.

Definition. *Let R be a ring and M, N be R-modules. A function $f : M \to N$ is an R-module homomorphism if it is a group homomorphism $M \to N$ and for all $r \in R$, $m \in M$*

$$f(rm) = rf(m)$$

An R-module homomorphism is sometimes referred to as a R-linear map. A bijective R-module homomorphism is an R-module isomorphism.

As usual we can define the submodules

$$\ker f = \{m \in M \mid f(m) = 0\}$$
$$\text{im } f = \{n \in N \mid \exists m \in M, f(m) = n\} = \{f(m) \mid m \in M\}$$

Example. A $\mathbb{Z}$-linear map of $\mathbb{Z}$-modules M, N is just a group homomorphism since they are also abelian groups.

Example. Let M be an R-module and $r \in R$ be some ring element. Then we can define an R-module homomorphism $f : M \to M$ by $f(m) = rm$

Definition. *$Hom_R(M, N)$ is the set of R-linear maps from $M \to N$ and is an R-module if R is commutative, otherwise it is only an abelian group. When $M = N$, we write $Hom_R(M, M) = End_R(M)$ which is a ring with identity known as the endomorphism ring. Naturally the elements of this ring are called endomorphisms.*

We now want to discuss quotients, it may be a bit surprising that there is no need for special submodules to quotient with. This is different from the case of groups and rings where we must consider normal subgroups and ideals respectively. This is because the underlying set of a module is necessarily an abelian group and thus all submodules are necessarily normal.

Definition. *Let R be a ring, M an R-module and N an R-submodule of M. The quotient module M/N is an abelian quotient group with the ring action*

$$r(m + N) = (rm) + N$$

where $m + N \in M/N$.

Let's verify that this action is well defined.

Proof. Suppose $m + N = m' + N$, this implies that

$$m - m' \in N \to r(m - m') \in N \to rm - rm' \in N$$

Thus $rm + N = rm' + N$ and so this is well defined. □

We can easily define a projection map

$$\pi : M \to M/N \qquad m \mapsto m + N$$

This is a R-module homomorphism with kernel N.
Now we can formulate the isomorphism theorems one more time, this time for modules. But first some definitions.

Definition. *Let A, B be submodules of some R-module M. The sum of A and B is*

$$A + B = \{a + b \mid a \in A, b \in B\} \leq M$$

Theorem 6.2 (The Isomorphism Theorems)**.** Here we list all four isomorphism theorems for modules

1. Let $f : M \to N$ be an R-linear map, then
$$M/\ker f \cong \text{im } f$$
2. Suppose $A, B \leq M$ are submodules, then
$$(A + B)/B \cong B/(A \cap B)$$
3. Suppose $A, B \leq M$ are submodules such that $A \subseteq B$, then
$$(M/A)/(B/A) \cong M/B$$
4. Let N be a submodule of M, there is a bijection
$$\{\text{submodules of } M/N\} \iff \{\text{submodules of } M \text{ containing } N\}$$

6.3 Direct Sums and Free Modules

Given some set of modules or submodules it is possible to generate new modules.

Definition. *Let M be an R-module and $N_1, \ldots, N_k$ be submodules of M. The sum of $N_1, \ldots N_k$ is the set*

$$N_1 + \cdots + N_k = \{a_1 + \cdots + a_k \mid a_i \in N_i\}$$

This sum of submodules is also a submodule.

Definition. *Let A be a subset of an R-module a, the submodule of M generated by A is the set*

$$RA = \left\{\sum r_i a_i \mid r_i \in R, a_i \in A\right\}$$

for a finite set $A = \{a_1, \ldots, a_n\}$ it is sometimes customary to write

$$RA = Ra_1 + \cdots Ra_n$$

If $N = RA$ is a submodule of M, then we say A is a generating set for N or that N is generated by A. N is finitely generated if A is finite. If N is generated by one element

$$N = Ra = \{ra \mid r \in R\}$$

then N is a cyclic submodule.

Most of these concepts should be familiar from previous discussions of groups and subgroups.

Definition. *Let $M_1, \ldots, M_k$ be a collection of R-modules. The direct sum of $M_1, \ldots, M_k$ is the set of tuples*

$$M_1 \times \cdots \times M_k = \{(m_1, \ldots, m_k) \mid m_i \in M_i\}$$

The is a module where addition and the ring action is component-wise.

This sometimes called the external direct sum and denoted

$$M_1 \oplus \cdots \oplus M_k$$

while the two are technically different they are only not the same in the case of infinite products/sums.

Proposition 6.3. Let $N_1, \ldots, N_k$ be submodules of an R-module M, then the following are equivalent

1. There is an R-module isomorphism

   $$\pi : N_1 + \cdots + N_k \cong N_1 \oplus \cdots \oplus N_k$$

   given by

   $$\pi(n_1, \ldots, n_k) = n_1 + \cdots + n_k$$

2. $N_j \cap (N_1 + \cdots + N_{j-1} + N_{j+1} + \cdots + N_k = 0$ for all j
3. Every $x \in N_1 + \cdots + N_k$ can be written uniquely in the form

   $$x = n_1 + \cdots + n_k$$

   where $n_i \in N_i$

Proof. Suppose (1) holds and assume that for some j, (2) is not true. Then there is some element n_j in that intersection which means

$$n_j = n_1 + \cdots + n_{j-1} + n_{j+1} + \cdots + n_k$$

But this is a nonzero element of the kernel which is a contradiction because π is an isomorphism thus (2) must be true.

Suppose (2) holds and the decomposition is not unique, that is

$$x = n_1 + \cdots + n_k = n_1' + \cdots + n_k'$$

then for each i we can write

$$(n_i - n_i') = (n_1 - n_1') + \cdots + (n_{j-1} + n_{j-1}') + (n_{j+1} + n_{j+1}') + \cdots + (n_k - n_k')$$

The right hand side is in the intersection and so must be zero, thus $n_i = n_i'$ and the sum is unique.

Now suppose (3) holds. π is clearly surjective and (3) just implies that it is also injective, thus it is an isomorphism. □

The proposition gives a condition in which the external direct sum is equal to what is known as the internal direct sum. For submodules $N_1, \ldots, N_k$, if they sum to M then we say that M is an internal direct sum of these submodules

$$M = N_1 + \cdots + N_k \to M = N_1 \oplus + \cdots + \oplus N_k$$

Note that the difference between an internal direct sum and an external direct sum are the type of the summands and their respective elements. For an internal direct sum, the summands are submodules and the elements are sums. For an external direct sum the summands are distinct modules and the elements are tuples. An internal direct sum is also an external direct sum if and only if the summands are pairwise disjoint.

Definition. *Let M be an R-module and $m_1, \ldots, m_n \in M$. These elements are linearly independent if*

$$r_1 m_1 + \cdots + r_n m_n = 0 \to r_1 = \cdots = r_n = 0$$

If $m_1, \ldots, m_n$ are linearly independent and generate M:

$$M = Rm_1 \oplus \cdots \oplus Rm_n$$

they they form a basis for M. In this case we say that M is a free module.

Since modules are a generalization of vector spaces it makes sense to have this concept as a generation of a vector space basis. There is a special map between free modules.

Proposition 6.4. Let M, N be free modules with basis $\{m_i\}, \{n_i\}$ respectively. Then there is a unique R-linear map $f : M \to N$ such that

$$f(m_i) = n_i \quad \forall i$$

Proof. We can write any element $m \in M$ as an R-linear combination

$$m = r_1 m_1 + \cdots + r_k m_k$$

where $\{m_i\}$ is a basis for M. Then by the properties of module homomorphisms

$$f(m) = r_1 f(m_1) + \cdots r_k f(m_k) = r_1 n_1 + \cdots r_k n_k$$

thus $f(m_i) = n_i$. It's easy to see that this map is unique. Suppose g is another such map, then

$$\begin{aligned} g(m) &= r_1 g(m_1) + \cdots r_k g(m_k) \\ &= r_1 n_1 + \cdots + r_k g_k \\ &= r_1 f(m_1) + \cdots + r_k f(m_k) \\ &= f(m) \end{aligned}$$

So $f = g$ and the map is unique. □

Thus there is a unique map between modules with equal sized basis which fixes basis elements. That is it maps a basis element of one module to a basis element of the other. This proposition gives a useful way to characterize free modules.

Proposition 6.5. IF M is a free module, then $M \cong R^{\oplus n}$ for some finite n.

Proof. Let $\{m_1, \ldots, m_n\}$ be a basis for M. By the previous proposition there is a unique map between this basis and the basis of $R^{\oplus n}$ (which is the set of tuples with a 1 in one coordinate and zeros elsewhere). This map is necessarily surjective because it is surjective when restricted only the basis of each module. Since the basis elements are linearly independent it is also injective (trivial kernel), making it an isomorphism. □

The proof requires that the basis elements be linearly independent which is of course how we define a basis. If there weren't then we'd only have a generating set and instead of M being a free module it would only be finitely generated. But this idea is also useful when characterizing finitely generated modules.

Proposition 6.6. A module m is finitely generated if there is a surjective map $R^{\oplus n} \to M$ for some finite n.

So if a module M is finitely generated there is a surjection

$$f : R^n \to M$$

(we will sometimes interchange R^n and $R^{\oplus n}$ for convenience). But this map has a kernel $\ker f \subseteq R^n$ which is a submodule of a finitely generated module. If the submodule of a finitely generated module is also finitely generated then we can continue this process

$$f_1 : R^{n_1} \to M_1$$

and so on for $M_2 = \ker f_1, M - 3, \ldots$. This iterative process is called the free resolution of a module M which is especially useful in homomological algebra. An R-module with this property means that R is what is called a Noetherian ring, to be discussed in more detail later.

Definition. *A ring R is Noetherian if, given a finitely generated R-module M and a submodule $M' \leq M$, then M' is also finitely generated.*

In particular if R is Noetherian then it is a finitely generated R-module generated by its multiplicative identity. This isn't the actual traditional definition but it will do for our purposes since we are discussing modules.

There is one part of this discussion which is missing. We mentioned that free modules and their bases are generalizations of vector spaces and their bases. It would make sense then to define the dimension of a module as the size of a basis, so why don't we? It turns out that each basis of an R-module M need not contain the same amount of elements unless R is commutative and nontrivial.

Proof. Let R be a commutative ring and M be a free R-module with two different bases

$$\{e_1, \ldots, e_n\}, \{f_1, \ldots, f_m\}$$

By a previous proposition this implies that $R^n \cong M \cong R^m$. Let I be a maximal ideal of R and fix an isomorphism $g : R^m \to R^n$, then there is an induced R-linear map

$$\overline{g} : R^m \to R^n / IR^n$$

consider some $x \in IR^m$

$$\overline{g}(x) = \overline{g}(i_1 m_1 + \cdots i_k m_k) = i_1 g(m_1) + \cdots + i_k g(m_k)$$

where $i_i \in I$ and $m_i \in R^m$. This implies that $\overline{g}(m) \in IR^n$ and thus is zero.

Conversely suppose $\overline{g}(m) = 0$ which implies $\overline{g}(m) \in IR^n$. We can write

$$\overline{g}(m) = i_1 n_1 + \cdots + \cdots i_k n_k$$

since g is an isomorphism we have

$$m = i_1 g^{-1}(n_1) + \cdots + i_k g^{-1}(n_k)$$

where $i_i \in I$ and $g^{-1}(n_i) \in R^m$, thus $m \in IR^m$. This implies that

$$\ker \overline{g} = IR^m$$

By the first isomorphism theorem we have

$$R^m / IR^m \cong R^n / IR^n$$

We can rewrite both sides as a single direct sum of a quotient

$$(R/I)^m \cong R^m / IR^m \cong R^n / IR^n \cong (R/I)^n$$

But R/I is a field since I is maximal and isomorphic vector spaces must have identical dimensions so $m = n$. □

Definition. *The rank of a free module over a nontrivial commutative ring is the size of any basis (since they all are the same size).*

6.4 The Fundamental Theorem of Modules over PIDs

The main goal of this section is to prove a structure theorem for finitely generated modules over Principle Ideal Domains. But first, some definitions...

Definition. *A left R-module is Noetherian if it satisfies the ascending chain condition on submodules, that is if there is a submodule chain*

$$M_1 \subseteq M_2 \subseteq M_3 \subseteq \cdots$$

then the chain terminates at some m. In other words for all $k > m$ we have $M_k = M_m$ and the chain becomes stationary. A ring R is Noetherian if it is Noetherian as a module over itself, that is there are no infinite increasing chains of ideals.

We gave another definition earlier of Noetherian rings, the two are equivalent.

Proposition 6.7. Let R be a ring and M a left R-module. The following are equivalent

1. M is Noetherian
2. Every nonempty set of submodules of M contains a maximal element under inclusion
3. Every submodule of M is finitely generated.

Proof. Suppose (1) holds and let $\{N_i\}_{i\in I}$ be a nonempty set of submodules of M indexed by I. For all i, let

$$S_i = \{j \in I \mid N_j \supset N_i\}$$

If N_i is maximal then $S_i = \varnothing$. Suppose that S_i is nonempty, then by the Axiom of Choice there is a choice function

$$f : \{S_i\}_{i\in I} \to I \qquad f(S_i) \in S_i$$

Choose some $i_0 \in I$ and set $i_1 = f(S_0), i_2 = f(S_1), \ldots$. We get the ascension

$$N_{i_0} \subset N_{i_1} \subset \cdots$$

which is a contradiction because M is Noetherian, thus S_i is empty and $\{N_i\}$ has a maximal element under inclusion.

Suppose (2) holds and consider a submodule N of M. Let X be the set of all finitely generated submodules of N which is nonempty because $\{0\} \in X$. Thus by (2) we can choose N' to be the maximal element of X. Suppose $N' \neq N$ and choose $n \in N \setminus N'$. Since N' is finitely generated by definition we can generated another submodule using N' and x. But this new submodule contains N' which is a contradiction because N' is maximal, thus $N = N'$ and is finitely generated.

Suppose (3) holds, consider an ascending submodule chain and define

$$M_1 \subset M_2 \subset \cdots \qquad N = \bigcup M_i$$

Note that N is a submodule of M and thus is finitely generated. Suppose N is generated by $\{n_1, \mathbb{N}_2, \ldots, n_k\}$ which means each n_i must lie in one of the submodule of the chain. Suppose $a_i \in M_{j_i}$ and define $m = \max\{j_1, \ldots, j_k\}$, then every $a_i \in M_m$ so that $N \subseteq M_m$. Thus $M_m = M_n = N$ for all $k \geq m$ and so M is Noetherian. □

Corollary 6.8. If R is a PID then it is Noetherian

Proof. Suppose R is a PID, the submodules of R are its ideals which are generated by a single element since they are principle. Thus R is Noetherian. □

The use fullness of Noetherian rings was touched upon briefly in an example earlier. If M is Noetherian then its submodules are finitely generated meaning that the submodules of a finitely generated module are also finitely generated. This is a stronger condition than requiring that M just be finitely generated. The final result we will prove before stating the structure theorem has to do with ranks of submodules.

Theorem 6.9. Let R be a PID and M a finitely generated R-module with a submodule N, then

1. N is free with $rank(N) \leq rank(M)$
2. There is a basis $e_1, \ldots, e_n$ and nonzero elements $a_1, \ldots, a_m \in R$ such that the set $\{a_1e_1, \ldots, a_me_m\}$ is a basis for N satisfying the division relation

$$a_1 \mid a_2 \mid \cdots \mid a_m$$

Proof. Assume $N \neq \{0\}$ because otherwise the proof is trivial. Consider some homomorphism $\phi : M \to R$. The image $\phi(N)$ is a submodule R (an ideal) and thus is finitely generated because R is PID, furthermore it is principle. Suppose $\phi(N) = (a_\phi)$ where $a_\phi \in R$.

Define the collection of submodules

$$\Sigma = \{\phi(N) \mid \phi \in Hom_R(M, R)\} = \{(a_\phi) \mid \phi \in Hom_R(M, R)\}$$

This collection is nonempty because we can simply take the trivial homomorphism which generates the ideal (0). Since R is Noetherian (since it's a PID), this collection must contain a maximal element. Let ν be a homomorphism such that $\nu(N) = (a_1)$ is that maximal element. We will also define $y \in N$ to be such that $\nu(y) = a_1$.

We want to show that a_1 is nonzero. Let $x_1, \ldots, x_n$ be a basis for M and let $\pi_i \in Hom_R(M, R)$ be the projection maps onto the ith coordinate. Since N is nontrivial, there must be some i for which $\pi_i(N) \neq 0$ which means that Σ contains more than just the trivial ideal (0). Since (a_1) is the maximal element of Σ it must be nonzero.

Now we will show that $a_1 \mid \phi(y)$ for every homomorphism $\phi : M \to R$. Let d be the generator for the ideal generated by $a_1, \phi(y)$, which means that it divides both a_1 and $\phi(y)$. This means we can write (using Euclid's algorithm)

$$d = r_1a_1 + r_2\phi(y) \qquad r_1, r_2 \in R$$

Now define the homomorphism $\psi = r_1\nu + r_2\phi$ and note

$$\psi(y) = r_1\nu(y) + r_2\phi(y) = r_1a_2 + r_2\phi(y) = d$$

This implies that $d \in \psi(N) \to (d) \subseteq \psi(N)$. Since d divides a_1 and the ideal (a_1) is maximal we have

$$(a_1) \subseteq (d) \subseteq \psi(N) \to (a_1) = (d) = \psi(N)$$

Since d also divides $\phi(y)$, we have $\phi(y) \in (a_1) \to a_1 \mid \phi(y)$.

We can apply this fact to the projection maps, in particular we have $a_1 \mid \pi_i(y)$ for any projection map π_i. Thus we can write $pi_i(y) = a_1b_i$ where $b_i \in R$. Define the sum

$$y_1 = \sum_{i=1}^{n} b_ix_i$$

and note that

$$a_1 y_1 = \sum_{i=1}^{n} a_i b_i x_i = \sum \pi_i(y) x_i = y$$

From this we get

$$a_1 = \nu(y) = \nu(a_1 y_1) = a_1 \nu(y_1) \to \nu(y_1) = 1$$

since $a_1 \neq 0$ and R is an integral domain.

This element y_1 can be taken as one element in a basis for M and $a_1 y_1$ is part of a basis for N. In particular

(a) $M = Ry_1 \oplus \ker \nu$

(b) $N = Ra_1 y_1 \oplus (N \cap \ker \nu)$

Now let's prove this. Consider some element $x \in M$ and note that we can write this as

$$x = \nu(x) y_1 + (x - \nu(x) y_1)$$

the second element satisfies

$$\begin{aligned} \nu(x - \nu(x) y_1) &= \nu(x) - \nu(x)\nu(y_1) \\ &= \nu(x) - \nu(x) = 0 \therefore \nu(x - \nu(x) y_1) \qquad \in \ker \nu \end{aligned}$$

This prove $M = Ry_1 + \ker \nu$. To see that this is direct, if $ry_1 \in \ker \nu$, then

$$\nu(ry_1) = r\nu(y_1) = r = 0$$

So the intersection is trivial and $M = Ry_1 \oplus \ker \nu$.

To prove (b) note that for any $x' \in N$, $\nu(x')$ is divisible by a_1 because a_1 generates $\nu(N)$ so we can write $\nu(x') = ba_1$ where $b \in R$. Thus

$$x' = \nu(x') y_1 + (x' - \nu(x') y_1) = ba_1 y_1 + (x' - ba_1 y_1)$$

Note that the second summand is an element of the kernel and also of N since $a_1 y_1 = y \in N$ (recall $\nu(y) = a_1$). As a consequence of (a) being a direct sum, this sum is also direct.

We can now start proving the theorem starting with showing that N is free of rank m. This will be done by induction on m, if $m = 0$ then N is torsion and $N = 0$ is free. Suppose this holds for some $m - 1$ and a submodule of rank $m - 1$ is free. For a direct sum $C = A \oplus B$, $rank(C) = rank(A) + rank(B)$. So if N has rank m then $N \cap \ker \nu$ has rank $m-1$ and is thus free by assumption. Since the sum is direct, we can adjoin $a_1 y_1$ to create a basis for N which means N is free of rank m. This proves that any submodule of M is free. Since $N \cap \ker \nu$ will has less than or equal rank to $\ker \nu$ we can also conclude that $rank(N) \leq rank(M)$.

We will prove (2) by induction on n, the rank of M. By (1), the submodule $\ker \nu$ is free of rank $n-1$ since the sum (a) is direct. By repeating the process in this proof with $M' = \ker \nu$

and $N' = N \cap \ker \nu$ we see that there is a basis $y_2, \ldots, y_n$ for $\ker \nu$ such that $a_2 y_2, \ldots, a_m y_m$ is a basis for $N \cap \ker \nu$. Since the sums are direct we can conclude that $y_1, \ldots, y_n$ is a basis for M and $a_1 y_1, \ldots, a_m y_m$ is a basis for N. Furthermore

$$a_2 \mid a_3 \mid \cdots \mid a_m$$

To complete the proof we must show that $a_1 \mid a_2$. Define a homomorphism

$$\varphi : M \to R \quad \varphi(y_1) = \varphi(y_2) = 1 \quad \varphi(y_i) = 0,\ i \geq 2$$

For this homomorphism we have

$$\varphi(a_1 y_1) = a_1 \to (a_1) \in \varphi(N)$$

Since (a_1) is maximal it must be that $\varphi(N) = (a_1)$. Since $\varphi(a_2 y_2) = a_2$, this also implies $a_2 \in (a_1)$ and so $a_1 \mid a_2$. □

This theorem shows that not only are submodules smaller (which seems obvious), but it is possible to construct a basis for a submodule from basis elements of the overlying module. Now for one more definition...

Definition. *The torsion submodule of a R-module M is the submodule*

$$Tor(M) = \{m \in M \mid rm = 0 \text{ for some nonzero } r \in R\}$$

If $Tor(M) = M$ we say that M is a torsion module and if $Tor(M) = \{0\}$ then M is torsion free. The annihilator of a submodule N of a module M is the ideal

$$Ann(N) = \{r \in R \mid rn = 0 \, \forall n \in N\}$$

Now we can state the structure theorem, which is also called the fundamental theorem of finitely generated modules over PIDs.

Theorem 6.10 (Fundamental Theorem; Invariant Factors)**.** Let R be a PID and M a finitely generated R-module, then

$$M \cong R^r \oplus R/(a_1) \oplus \cdots \oplus R/(a_m)$$

where $r \geq 0$ and $a_i \in R$ are nonzero non-unit elements such that

$$a_1 \mid a_2 \mid \cdots a_m$$

Proof. Suppose M is generated by $x_1, \ldots, x_n$ and let R^n be the free R-module of rank n with a basis $b_1, \ldots, b_n$. We can define a surjective homomorphism

$$\pi : R^n \to M \qquad \pi(b_i) = x_i$$

since x_i generate M, thus by the first isomorphism theorem

$$R^n / \ker \pi \cong M$$

By the previous theorem there is some basis $y_1, \ldots, y_n$ for R^n and elements $a_1, \ldots, a_m$ such that $a_1 y_1, \ldots, a_m y_m$ are a basis for $\ker \pi$. The elements a_i satisfy the division relations $a_i \mid a_{i+1}$. Thus we can write the quotient as

$$M \cong (Ry_1 \oplus \cdots \oplus Ry_n)/(Ra_1 y_1 \oplus \cdots \oplus Ra_m y_m)$$

To identify this quotient consider the naturally surjective homomorphism

$$Ry_1 \oplus \cdots \oplus Ry_n \to R/(a_1) \oplus \cdots \oplus R/(a_m) \oplus R^{n-m}$$

given by

$$(r_1 y_1, \ldots, r_n y_n) \mapsto (r_1 \mod a_1, \ldots, r_m \mod a_m, r_{m+1}, \ldots, r_n)$$

The kernel is the set of elements where a_i divides r_i with zeroes elsewhere, which is just the direct sum $Ra_1 y_1 \oplus \cdots \oplus Ra_m y_m$. Thus by the first isomorphism theorem again we can rewrite

$$M \cong R/(a_1) \oplus \cdots \oplus R/(a_m) \oplus R^{n-m}$$

Identifying $r = n - m$ and noting that if a is a unit then $R/(a) = 0$ we get the desired decomposition of the theorem. □

Definition. *The integer r is called the free rank or Betti number of M, and the elements a_i are called the invariant factors of M.*

By examining the theorem we see that every finitely generated module over a PID is a direct sum of a free module and a torsion module, hence the name structure theorem. Specifically

$$Tor(M) \cong R/(a_1) \oplus \cdots \oplus R/(a_m)$$

thus M is a torsion module if its free rank is zero and it is torsion free if and only if it is a free module. There exists another decomposition in which we use the Chinese Remainder Theorem to rewrite the cyclic modules in way that makes their annihilators either (0) or generated by a prime power.

Theorem 6.11 (Fundamental Theorem; Elementary Divisors)**.** Let R be a PID and M a finitely generated R-module, then

$$M \cong R^r \oplus R/(p_1^{\alpha_1}) \oplus \cdots \oplus R/(p_s^{\alpha_s})$$

where $r \geq 0$ and $p_i^{\alpha_i}$ are positive powers of (not necessarily distinct) primes in R.

Proof. Let a be a nonzero element in R. Since R is also a UFD we have the factorization

$$a = u p_1^{\alpha_1} \cdots p_s^{\alpha_s}$$

where u is a unit and p_i are distinct primes. This factorization is unique up to units so the ideals $(p_i^{\alpha_i})$ are uniquely defined. The GCD of any two primes is 1 so for $i \neq j$ we have

$$(p_i^{\alpha_i}) + (p_j^{\alpha_j}) = R$$

Furthermore since a is the least common multiple, the intersection of all these ideals is (a). Thus the Chinese Remainder Theorem gives

$$R/(a) \cong R/(p_1^{\alpha_1}) \oplus \cdots \oplus p_s^{\alpha_s}$$

Thus we can rewrite all the cyclic summands $R/(a_i)$ in this form which gives us the desired decomposition of the theorem. □

Definition. *The prime powers $p_i^{\alpha_i}$ are called the elementary divisors of M.*

We now want to investigate the uniqueness of these decompositions, but first we prove a useful lemma.

Lemma 6.12. Let R be a PID and p be prime in R (and thus maximal) and denote the field $F = R/(p)$.

1. Suppose $M = R^r$, then $M/pM \cong F^r$
2. Let $M = R/(a)$ for some nonzero $a \in R$, then

$$M/pM \cong \begin{cases} F & \text{if } p \mid a \text{ in } R \\ 0 & \text{otherwise} \end{cases}$$

3. Let

$$N = R/(a_1) \oplus \cdots \oplus R/(a_k)$$

and suppose $p \mid a_i$ for all i, then

$$M/pM \cong F^k$$

Proof. 1) There is a natural surjective map

$$R^r \to (R/(p))^r \quad (a_1, \ldots, a_r) \mapsto (a_1 \mod p, \ldots, a_r \mod p)$$

The kernel of this is the elements divisible by p, which is pR^r, thus by the first isomorphism theorem

$$R^r/pR^r \cong (R/(p))^r \to M/pM \cong F^r$$

2) Note that $p(R/(a)) = ((p) + (a))/(a)$. Consider the ideal $(p) + (a)$, since we are in a PID this is generated by a single element, namely the GCD of p and a. If $p \mid a$ then $(p)+(a) = (p)$, otherwise it is $(1) = R$. Thus

$$pM = \begin{cases} (p)/(a) & \text{if } p \mid a \text{ in } R \\ R/(a) = M & \text{otherwise} \end{cases}$$

If we form the quotient we will get (using the third isomorphism theorem)

$$M/pM \cong \begin{cases} (R/(a))/((p)/(a)) \cong R/(p) = F & \text{if } p \mid a \text{ in } R \\ M/M = 0 & \text{otherwise} \end{cases}$$

3) Note that there is a natural surjective map

$$R/(a_1) \oplus \cdots \oplus R/(a_k) \to (R/(a_1))/p(R/(a_1)) \oplus \cdots \oplus (R/(a_k))/p(R/(a_k))$$

given by

$$(r_1 + (a_1), \ldots, r_k + (a_k)) \mapsto ((r_1 \mod p) + (a_1), \ldots, (r_k \mod p) + (a_k))$$

with kernel pM. Thus we get the quotient

$$M/pM \cong (R/(a_1))/p(R/(a_1)) \oplus \cdots \oplus (R/(a_k))/p(R/(a_k)) \cong F \oplus \cdots \oplus F = F^k$$

as required. □

Theorem 6.13 (Fundamental Theorem; Uniqueness)**.** Let R be a PID

1. Two finitely generated R-modules are isomorphic if they have the same free rank and identical invariant factors.
2. Two finitely generated R-modules are isomorphic if they have the same free rank and identical elementary divisors.

This theorem tells us that the decompositions are unique, at least up to isomorphism. The next corollary will give us a way to convert between invariant factors and elementary divisors.

Corollary 6.14. Let R be a PID and M a finitely generated R-module

1. The elementary divisors of M are the prime power factors of the invariant factors.
2. The largest invariant factor of M is the product of the largest of distinct prime powers among the elementary divisors of M. The next largest factor is the product of the next largest of distinct prime powers among the remaining divisors and so on.

Proof. These procedures gives us a list of elementary divisors (or invariant factors) and since they are unique, they must be *the* elementary divisors (or invariant factors). □

This procedure is illustrated in the case of finitely generated abelian groups which we showed can be considered as $\mathbb{Z}$-modules. Thus applying the structure theorem in the case of $\mathbb{Z}$-modules will gives us the Fundamental Theorem of Finitely Generated Abelian Groups discussed earlier.

6.5 Tensor Products of Modules

The tensor product of two modules is a construction that allows for the creation of a new module in which the elements are products.

Definition. *Let M be a right R-module and N a left R-module. The tensor product $M \otimes_R N$ is the abelian group generated by*

$$\{m \otimes n \mid m \in M, n \in N\}$$

with the relations

$$(m_1 + m_2) \otimes n = m_1 \otimes n + m_2 \otimes n$$
$$m \otimes (n_1 + n_2) = m \otimes n_1 + m \otimes n_2$$
$$mr \otimes n = m \otimes rn$$

The first two relations imply bilinearity and the last implies R-linearity. The elements of a tensor product are called tensors.

Example. Consider the tensor product $\mathbb{Z}/2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/3\mathbb{Z}$ which is spanned by

$$\{0 \otimes 0, 0 \otimes 1, 0 \otimes 2, 1 \otimes 0, 1 \otimes 1, 1 \otimes 2\}$$

We can use the tensor relations to simplify this basis, for instance $0 \times a = 1 \otimes a - 1 \otimes a = 0$ which eliminates half of these. A similar argument shows that $1 \otimes 0 = 0$ Furthermore note that

$$1 \otimes 2 = 1 \otimes 1 + 1 \otimes 1$$

so this tensor will be generated by $1 \otimes 1$, but

$$\begin{aligned} 1 \otimes 1 &= 1 \otimes 3 - 1 \otimes 2 \\ &= 1 \otimes 0 - 1 \otimes 2 \\ &= 1 \otimes 0 - 2 \otimes 1 \\ &= 1 \otimes 0 - 0 \otimes 1 = 0 \end{aligned}$$

Thus this tensor product is just zero, $\mathbb{Z}/2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/3\mathbb{Z} = 0$

Example. For a nontrivial example consider free modules over a field k, $V = k^n$ and $W = k^n$ and suppose that $\{e_i\}$ and $\{f_i\}$ are their respective basis. Then $V \otimes_k W$ is spanned by

$$\{v \otimes w \mid v \in V, w \in W\}$$

We can write elements in V and W in terms of their respective bases as

$$v = \sum a_i e_i \qquad w = \sum b_i f_i \qquad a_i, b_i \in k$$

Consider some arbitrary tensor $v \otimes w$ and note

$$v \otimes w = \left(\sum a_i e_i\right) \left(\sum b_j f_j\right) = \sum a_i b_j (e_i \otimes f_j)$$

So this vector space is actually spanned by $\{e_i \otimes f_j\}$ so the tensor product is just a k-vector space of dimension $\dim V \cdot \dim W = nm$.

But we defined the tensor product to be an abelian group so why is it a vector space here? It turns out that if certain conditions are met then the tensor product of modules still has a module structure.

Proposition 6.15. If R is a commutative ring and M, N R-modules then the tensor product $M \otimes_R N$ carries the structure of an R-module under the action

$$r(m \otimes n) = (rm) \otimes n = m \otimes (rn)$$

Definition. *A (R, S)-bimodule is an abelian group M which is a left R-module and right S-module such that*

$$(rm)s = r(ms)$$

The tensor protect can also preserve a bimodule structure under certain conditions.

Proposition 6.16. If M is a (R, S)-bimodule and N a (S, T)-bimodule, then $M \otimes_S N$ is a (R, T)-module under the action

$$r(m \otimes n)t = (rm) \otimes (nt)$$

Example. A ring R is a bimodule for itself. If $\phi : R \to S$ is a ring homomorphism then S is a (R, S)-bimodule under

$$rss' = \phi(r)ss'$$

Thus if M is a right R-module (so a $(\mathbb{Z}, R)$-bimodule), then $M \otimes_R S$ is a right S-module.

We find from the previous example that $- \otimes_R S$ is a functor from right R-modules to right S-modules and likewise for $S \otimes_R -$ a functor of left modules. We will investigate the functor more later.

From the last few examples we can see that it is often tedious to manipulate tensors using the relations to determine if they are nonzero. There is an easier way.

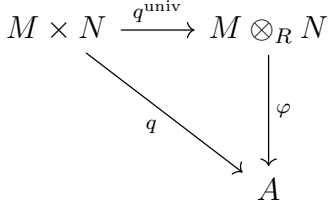
Definition. *Let M be a left R-module, N a right R-module, and A an abelian group. A R-bilinear map is a map $q : M \times N \to A$ such that*

$$q(m_1 + m_2, n) = q(m_1, n) + q(m_2, n)$$
$$q(m, n_1 + n_2) = q(m, n_1) + q(m, n_2)$$
$$q(mr, n) = q(m, rn)$$

By definition there is a natural bilinear map

$$q : M \times N \to M \otimes N \qquad q(m, n) = m \otimes n$$

In fact for every abliean group A and bilinear map $q : M \times N \to A$ there is a unique factorisation



where $\varphi : M \otimes_R N \to A$ is a homomorphism of abelian groups. We can use this to show that certain elements are nonzero, namely if $q(m, n) \neq 0$ then $m \otimes n \neq 0$. This is known as the universal property for tensor products.

Example. Consider the tensor product $\mathbb{Z}/3\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/3\mathbb{Z}$. We can define

$$q : \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z} \to \mathbb{Z}/3\mathbb{Z} \qquad q(m,n) = mn$$

Thus $q(m,n) = 0$ only if $m = 0$ or $n = 0$. For nonzero $a, b \in \mathbb{Z}$ then

$$a \otimes b = ab(1 \otimes 1)$$

which seems to imply that the tensor product is cyclic

$$\mathbb{Z}/3\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/3\mathbb{Z} \cong \mathbb{Z}/3\mathbb{Z}$$

In general we can construct a bilinear map

$$\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/d\mathbb{Z}$$

where $d = (m,n)$ is the GCD. Thus also provides an easier way to show an earlier example.

Example. Recall that $k^n \otimes_k k^m$ is a k-vector space spanned by $\{e_i \otimes f_j\}$. We can construct a bilinear map

$$q : k^n \times k^m \to M_{m\times n}(k) \qquad q(v,w) = vw^T$$

Thus if there is some zero tensor then by the universal property

$$\sum a_{ij} e_i \times f_j = \sum a_{ij} E_{ij} = 0 \to a_{ij} = 0 \, \forall i, j$$

where E_{ij} is the matrix with a 1 at row i column j and zeroes elsewhere. The last implication follows because the set $\{E_{ij}\}$ is a basis for $M_{m\times n}(k)$.

This last example shows that while $M \otimes N$ is spanned by $m \otimes n$ (called pure tensors), not every element can be expressed in this form. For instance in the previous example this is because not every matrix can be expressed in the form vw^T, in other words not every matrix is rank 1.

Example. If A, B, C are commutative rings and there exists a ring homomorphism $\varphi : C \to A$ and $\psi : C \to B$, then A, B are both C-modules and $A \otimes_C B$ is a commutative ring.

Example. The tensor product $k[x] \otimes_k k[y]$ has basis $x^i \otimes y^j = x^i y^j$ and the natural ring structure of $k[x,y]$. A pure tensor is of the form $f(x)g(y)$ so we see that in this tensor product not every element can be expressed as a pure tensor ($x^2 + y^2$ for example).

We will now examine the tensor product as a functor in more detail, a discussion which will continue into the next section. If M, N are right R-modules, P a left R-module, and $f : M \to N$ a R-linear map, then there is a map

$$f \otimes P : M \otimes_R P \to N \otimes_R P \qquad f \otimes P(m \otimes p) = f(m) \otimes P$$

Note that this is well defined because

$$\begin{aligned} f \otimes P(mr \otimes p) &= f(rm) \otimes p \\ &= f(m)r \otimes p \\ &= f(m) \otimes rp \\ &= f \otimes P(m \otimes rp) \end{aligned}$$

Thus we see that $- \otimes P$ is a functor from right R-modules to abelian groups. When R is commutative then this is a functor of R-modules.

Proposition 6.17. If $f : M \to N$ is a surjective homomorphism, then

$$f \otimes P : M \otimes_R P \to N \otimes_R P$$

is also a surjective map.

Proof. We only need to show that $n \otimes p = \text{im } (f \otimes P)$ for all $n \in N$ since the pure tensors span the tensor product $N \otimes_R P$. Since f is surjective there exists some m such that $f(m) = n$ which means

$$f \otimes P(m \otimes p) = f(m) \otimes p = n \otimes p$$

thus $f \otimes P$ is surjective. □

Note that the tensor map need not preserve injectivity as we will see in the next example.

Example. Let $R = \mathbb{Z}$ and $P = \mathbb{Z}/p\mathbb{Z}$ for a prime p. Consider the map

$$f : \mathbb{Z} \to \mathbb{Z} \qquad x \mapsto px$$

which induces the tensor map

$$f \otimes \mathbb{Z}/p\mathbb{Z} : \mathbb{Z} \otimes \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z} \otimes \mathbb{Z}/p\mathbb{Z}$$

and note that

$$1 \otimes 1 \mapsto p \otimes 1 = 1 \otimes p = 1 \otimes 0 = 0$$

so this is clearly not injective

Example. Let A be a commutative ring, I an ideal, and M an A-module. Then

$$M \otimes_A A/I \cong M/IM$$

We will prove this by constructing maps in both directions.

Proof. Suppose we define the map $\phi(m) = m \otimes 1$ for $m \in M$. Note that by the universal property we have a canonical map $(m, 1) \mapsto m \otimes 1 \in M \otimes A/I$. Suppose $a \in I$ and $m = a \cdot m'$, then

$$\begin{aligned} \phi(m) &= am' \otimes 1 \\ &= m'a \otimes 1 \\ &= m' \otimes a \\ &= m' \otimes 0 = 0 \end{aligned}$$

In other words $IM \mapsto 0$, thus ϕ is a well defined map from $M/IM \to M \otimes A/I$.

For the other direction we cannot simply just say what $\psi(m \otimes a)$ is because m, a are not well defined, onloy the tensor is. Instead we will use the universwal property. Define the map

$$M \times A/I \to M/IM \qquad (m, \overline{a}) \mapsto \overline{am}$$

To verify that this is A-balanced note

$$(mb, \overline{a}) \mapsto \overline{amb}$$
$$(m, b\overline{a}) = (m, \overline{ba}) \mapsto \overline{bam}$$

Thus there exists a unique map $\psi : M \otimes A/I \to M/IM$ such that $\psi(m \otimes \overline{a}) = \overline{ma}$.

To complete this proof we just need to show that $\phi^{-1} = \psi$.

$$\overline{m} \mapsto \overline{m} \otimes 1 \mapsto \overline{1m} = \overline{m}$$
$$m \otimes \overline{a} \mapsto \overline{am} \mapsto am \otimes 1 = m \otimes \overline{a}$$

Thus these are indeed inverses and we have the desired isomorphism. □

Using the same method as above we can prove the following useful properties.

Proposition 6.18. Let M, N, P be A-modules, then there exists unique isomorphisms

1. $M \otimes N \to N \otimes M$
2. $(M \otimes N) \otimes P \to M \otimes (N \otimes P) \to M \otimes N \otimes P$
3. $(M \oplus N) \otimes P \to (M \otimes P) \oplus (N \otimes P)$
4. $A \otimes M \to M$

such that, respectively,

1. $x \otimes y \mapsto y \otimes x$
2. $(x \otimes y) \otimes z \mapsto x \otimes (y \otimes z) \mapsto x \otimes y \otimes z$
3. $(x, y) \otimes z \mapsto (x \otimes z, y \otimes z)$
4. $a \otimes x \mapsto ax$

so in a sense the above isomorphisms form identities.

The above properties all work if we tensor over a single ring, for tensoring over multiple rings we have another fundamental isomorphism. If we have three modules ${}_AM_B, {}_BN_C, {}_CP_D$, then

$$(M \otimes_B N) \otimes_C P \cong M \otimes_B (N \otimes_C P)$$

For the most part these properties are similar to distributing addition and multiplication.

Similar to how ideals can be formed by extending and contracting through a homomorphisms, modules can be obtaining by restricting and extending scalars. Let $f : A \to B$ be a homomorphism of rings and N a B-module. We can give N an A-module structure using the action

$$ax = f(a)x$$

This A-module structure is said to be obtained from N by restriction of scalars, note that we can also define an A-module structure on the ring B the same way.

Proposition 6.19. Suppose N is finitely generated as a B-module and B is finitely generated as an A-module, then N is finitely generated as an A-module.

Proof. Let $y_1, \ldots, y_n$ generate N over B and $x_1, \ldots, x_m$ generate B as an A-module, then the mn products $x_i y_j$ generate N over A. □

Now suppose M is an A-module, since B can be viewed as an A-module we can create the A-module $M_B = B \otimes_A M$. In fact M_B has a B-module structure with the action

$$b(b' \otimes x) = bb' \otimes x$$

We say that M_B was obtained from M by extension of scalars.

Proposition 6.20. If M is finitely generated as an A-module, then M_B is finitely generated as a B-module.

Proof. Let $x_1, \ldots, x_n$ be generators for M over A, then the pure tensors $1 \otimes x_i$ generate M_B over B. □

6.6 Exact Sequences

Recall in group theory that we can break up groups

$$N \trianglelefteq G \to G/N$$

using a central series. We can do something similar for modules.

Definition. *An exact sequence of left (right) R-modules is a chain of R-module morphisms*

$$M_1 \xrightarrow{d_1} M_2 \xrightarrow{d_2} M_3 \longrightarrow \cdots \longrightarrow M_{N-1} \xrightarrow{d_{N-1}} M_n$$

such that

1. $d_{i+1} \circ d_i : M_i \to M_{i+1}$ *is the zero map (chain is a complex). This implies*

$$\ker d_{i+1} \supset im\ d_i$$

2. $\ker d_{i+1} = im\ d_i$ *(chain is exact)*

For a non exact complex, $\ker d_{i+1}/\text{im } d_i$ is called the homomology of the complex at M_{i+1}. The existence of an exact sequence has several useful implications.

Proposition 6.21. Suppose we have an exact sequence

$$0 \xrightarrow{d_0} M_1 \xrightarrow{d_1} M_2 \longrightarrow \cdots \xrightarrow{d_{N-1}} M_N \xrightarrow{d_N} 0$$

Exactness at M_1 implies

$$\ker d_1 = \text{im } d_0 = 0 \to d_1 \text{ is injective}$$

Exactness at M_N implies

$$M_N = \ker d_N = \text{im } d_{N-1} \to d_{N-1} \text{ is surjective}$$

Proof. Clear from definition □

This implies that if we have an exact sequence

$$0 \longrightarrow A \xrightarrow{\varphi} B \longrightarrow 0$$

Then $\varphi : A \to B$ is an isomorphism. The next simplest case involves three nontrivial modules.

Definition. *A short exact sequence is*

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

Exactness implies that f is injective, g is surjective and

$$\ker g = im\ f = A$$

which implies (by isomorphism theorems)

$$im\ g = C \cong B/A = B/\ker g$$

Example. Consider a homomorphism $\varphi : B \to C$, there is a natural exact sequence

$$0 \longrightarrow \ker \varphi \xrightarrow{\iota} B \xrightarrow{\varphi} \text{im } \varphi \longrightarrow 0$$

In particular if φ is surjective then the sequence $B \to C$ can be extended to a short exact sequence with $A = \ker \varphi$.

We will now examine exact sequences under tensor products.

Proposition 6.22. Let A, B, C be right R-modules and P a left R-module. Consider the exact sequence

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

then the sequence

$$0 \longrightarrow A \otimes P \xrightarrow{f\otimes P} B \otimes P \xrightarrow{g\otimes P} C \otimes P \longrightarrow 0$$

is exact at $B \otimes P$ and $C \otimes P$, in other words $g \otimes P$ is surjective.

Proof. Since $g : B \to C$ is surjective due to exactness at C, the map $g\otimes P : B\otimes P \to C\otimes P$ is also surjective which means the sequences is exact at $C \otimes P$. We now need to show exactness at $B \otimes P$. Let $D = \text{im } f \otimes P$ and so $D \subseteq \ker g \otimes P$. If $D = \ker g \otimes P$ then the map

$$g \otimes P : B \otimes P/D \to C \otimes P$$

is an isomorphism (since we already know it is surjective).

We show that this is an isomorphism by constructing an inverse ψ and stating where the pure tensor $c \otimes p$ gets mapped. Because g is surjective, let $b \in B$ be the element such that $g(b) = c$ and define $\psi(c \otimes P) = b \otimes p$.

Suppose $g(b') = c$, then

$$b \otimes p - b' \otimes p = (b - b') \otimes p$$

But $b - b' \in \ker g = \text{im } f$ due to exactness at B so there exists an element $a \in A$ such that $f(a) = b - b'$ which implies

$$(b - b') \otimes p = f \otimes P(a \otimes p) \in D$$

Thus we've shown that $D = \text{im } f \otimes P = \ker g \otimes P$ and so $g \otimes P$ is injective. □

A functor which is only exact at certain parts of the sequence has a special name.

Definition. *Consider a functor F and a short exact sequence*

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

If F is exact at B, C then it is right exact and we have the exact sequence

$$F_A \xrightarrow{F_f} F_B \xrightarrow{F_g} F_C \longrightarrow 0$$

If F is exact at A, B then it is left exact and we have the exact sequence

$$0 \longrightarrow F_A \xrightarrow{F_f} F_B \xrightarrow{F_g} F_C$$

If F is exact at A, B, C then it is referred to as just an exact functor.

Thus we see that $- \otimes P$ is a right exact functor from R-modules to abelian groups.

Exact sequences have useful applications. For instance let R be a ring, M a right R-module, and I a left ideals. Then we can construct a short exact sequence of left R-modules

$$0 \longrightarrow I \xrightarrow{\iota} R \xrightarrow{\pi} R/I \longrightarrow 0$$

Where ι is the inclusion map and π is the structure map. Tensoring gives the sequence

$$0 \longrightarrow M \otimes I \longrightarrow M \otimes R \cong M \longrightarrow M \otimes (R/I) \longrightarrow 0$$

Let $m \otimes i$ be a pure tensor in $|M \otimes I$ which has an imagine in $M \otimes R$ of $m \otimes i$ under inclusion. But this is just mi under the isomorphism $M \otimes R = M$ given by

$$m \otimes r = rm \otimes 1 \cong mr$$

So the image of $M \otimes I$ in M is the subgroup

$$MI = span\,\{mi \mid m \in M, i \in I\} \subset M$$

But this sequence is exact which means

$$M \otimes (R/I) \cong M \otimes R/\text{im } (M \otimes I) = M/MI$$

Example. The isomorphism we just showed can also be used to shed light on some tensor products. If A is a $\mathbb{Z}$-module (so an abelian group), then

$$A \otimes \mathbb{Z}/m\mathbb{Z} \cong A/mA$$

Using this we find

$$\mathbb{Z}/2\mathbb{Z} \otimes \mathbb{Z}/3\mathbb{Z} = (\mathbb{Z}/2\mathbb{Z})/3(\mathbb{Z}/2\mathbb{Z}) = (\mathbb{Z}/2\mathbb{Z})/(\mathbb{Z}/2\mathbb{Z}) = 0$$
$$\mathbb{Z}/3\mathbb{Z} \otimes \mathbb{Z}/3\mathbb{Z} = (\mathbb{Z}/3\mathbb{Z})/3(\mathbb{Z}/3\mathbb{Z}) = (\mathbb{Z}/3\mathbb{Z})/0 \cong \mathbb{Z}/3\mathbb{Z}$$

It's interesting to note that both $M \otimes_R -$ and $- \otimes_R P$ are right exact. In general the functor is not completely exact but it can happen in some cases. For instance $- \otimes_R R$ and $- \otimes R^n$ are both exact functors. The functor $- \otimes_z \mathbb{Q}$ is also exact. The exactness of the tensor functor defines a special type of module.

Definition. *When the functor $- \otimes P$ is exact then P is a flat module.*

As a final useful property, exact sequences can help determine ranks.

Proposition 6.23. Consider the exact sequence of abelian groups ($\mathbb{Z}$-modules)

$$0 \longrightarrow A \longrightarrow B \longrightarrow C \longrightarrow 0$$

The ranks satisfy

$$rank(B) = rank(A) + rank(C)$$

Proof. Using the fundamental theorem for finitely generated abelian groups

$$A \otimes_{\mathbb{Z}} \mathbb{Q} = (\mathbb{Z}^k \oplus \mathbb{Z}/m_1\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/m_s\mathbb{Z}) \otimes \mathbb{Q}$$

Note that $(M \oplus N) \otimes P = (M \otimes P) \oplus (N \otimes P)$. Using this fact we get

$$A \otimes_z \mathbb{Q} = \mathbb{Z}^k \otimes \mathbb{Q} = \mathbb{Q}^k$$

since $\mathbb{Z}/m\mathbb{Z} \otimes \mathbb{Q} = \mathbb{Q}/m\mathbb{Q} = 0$. Thus we can associate the rank with

$$rank(A) = \dim A \otimes_{\mathbb{Z}} \mathbb{Q}$$

The functor $- \otimes_{\mathbb{Z}} \mathbb{Q}$ is exact so we have the exact sequence

$$0 \longrightarrow A \otimes \mathbb{Q} \longrightarrow B \otimes \mathbb{Q} \longrightarrow C \otimes \mathbb{Q} \longrightarrow 0$$

which implies the isomorphism

$$C \otimes \mathbb{Q} = B \otimes \mathbb{Q}/A \otimes \mathbb{Q}$$

Thus the dimensions satisfy

$$\dim B \otimes_{\mathbb{Z}} \mathbb{Q} = \dim A \otimes_{\mathbb{Z}} \mathbb{Q} + \dim C \otimes_{\mathbb{Z}} \mathbb{Q}$$

which proves the desired rank relation. □

There are two more types of modules which we will go over briefly.

Definition. *An A-module P is projective if $Hom_A(P, -)$ is exact*

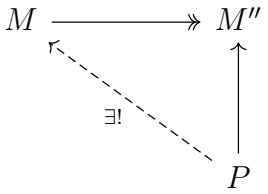
Thus if a module is projective and we have the exact sequence

$$0 \longrightarrow M' \longrightarrow M \longrightarrow M'' \longrightarrow 0$$

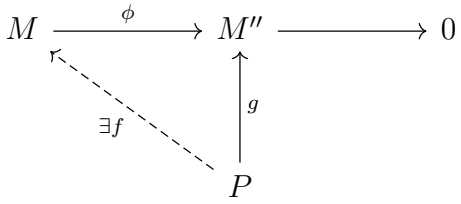
then we also have the exact sequence

$$0 \longrightarrow Hom(P, M') \longrightarrow Hom(P, M) \longrightarrow Hom(P, M'') \longrightarrow 0$$

An equivalent definition for projective modules is if $M \to M''$ is surjective, then so is $Hom(P, M) \to Hom(P, M'')$. We can represent this in a diagram: P is projective if and only if[4]



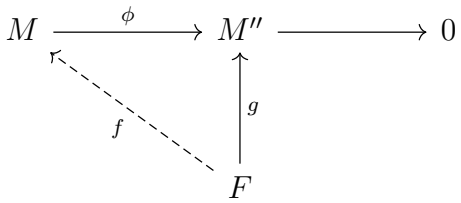
In other words if we have a surjective map $\phi : M \to M''$ and another map $g : P \to M''$, then there exists a unique map $f : P \to M$ such that g factor through f, in diagram form this is



Projective modules are intrinsically related to free modules as we will see from the following example and propositions.

Example. A free module is projective.

Proof. Let F be a projective module and suppose we have the diagram



Let S be a basis for F, for all $s \in S$ consider $g(s) \in M''$. Since ϕ is surjective there exists a $m_s \in M$ such that $\phi(m_s) = g(s)$. Now since F is free, we can map S to M any way we want to get a module homomoprhism $f(s) = m_s$. □

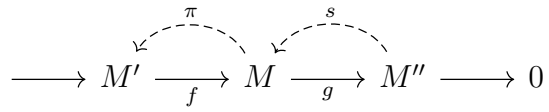
[4]the double arrows indicates a surjective map

In fact if $F = P \oplus Q$ is free, then both direct summands P and Q are projective modules. We will show that if a module is projective, then there exists another modules such that their direct sum is free.

Lemma 6.24. Consider a short exact sequence

$$0 \longrightarrow M' \xrightarrow[f]{} M \xrightarrow[g]{} M'' \longrightarrow 0$$

A splitting of f is a map $\pi : M \to M'$ such that $\pi \circ f = id$, we can construct a splitting for g in a similar way (see the diagram below)



The following are true about splitting:

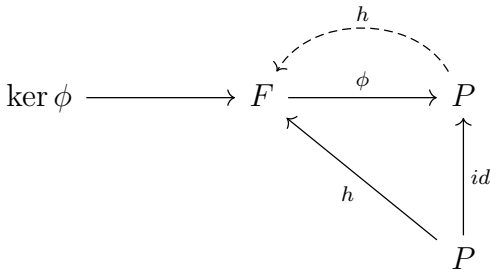
1. π exists if and only if s exists
2. If either π or s exists, then $M \cong M' \oplus M''$

Theorem 6.25. If P is projective, then it is a direct summand of a free module

Proof. Let $S \subseteq P$ be a set of generators (note that we can also just take S to be all of P). Let F be the free A-module on the set S, this gives us a map

$$\phi : F \to P \qquad s \in F \mapsto s \in P$$

Since S generates P, this map is onto, but since this P is projective we have some $h : P \to F$ such that $\phi \circ h = id$ (see diagram)



In other words h splits the sequence and thus

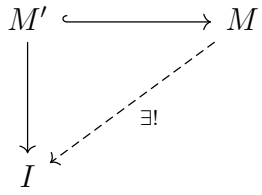
$$F \cong P \oplus \ker \phi$$

□

A projective module P will also be flat, the converse is only true if P is a finitely generated flat A-module with A a Noetherian ring. The dual notion of a projective module is the injective module.

Definition. *A module I is injective if $Hom_A(-, I)$ is exact.*

The equivalent commutative diagram is[5]



The relationship between injective and projective modules can be seen with the following two theorems.

Theorem 6.26. Let M by any A-module, then there exists an injective map

$$P \to M \to 0$$

from some projective module P.

Theorem 6.27 (Dual)**.** Let M by any A-module, then there exists an injective map

$$0 \to M \to I$$

to some injective module I.

7 Topics in Commutative Algebra

7.1 Semisimple Rings

Recall that a group ring is the set of sums

$$R[G] = \left\{ \sum a_g g \mid a_g \in R, g \in G \right\}$$

Generally we consider the case where R is commutative and G is finite. Recall that a $\mathbb{C}[G]$-module is the same as a complex representation of G.

Example. Consider the group ring $\mathbb{Z}[\mathbb{Z}/4\mathbb{Z}]$ which is the set of sums

$$a_0 0 + a_1 1 + a_2 2 + a_3 3 \quad a_i \in \mathbb{Z}$$

We can represent the elements of $\mathbb{Z}/4\mathbb{Z}$ as $\{1, x, x^2, x^3\}$

$$\mathbb{Z}[\mathbb{Z}/4\mathbb{Z}] = \left\{ a_0 + a_1 x + a_2 x^2 + a_3 x^3 \mid a_i \in \mathbb{Z} \right\} = \mathbb{Z}[x]/(x^4 - 1)$$

Note that this is not an integral domain since $(x^2 + 1)(x^2 - 1) = 0$. We can break up this ring using a map

$$\mathbb{Z}[x]/(x^4 - 1) \to \mathbb{Z}[x]/(x^2 + 1) = \mathbb{Z}[i]$$

[5] curved arrow is an injective map

since $(x^2+1) \subset (x^4-1)$. We can think of this as a homomorphism

$$\mathbb{Z}[\mathbb{Z}/4\mathbb{Z}] \to \mathbb{Z}[i] \qquad x \mapsto i$$

Similarly we can send $x \mapsto 1$ and get $\mathbb{Z}[\mathbb{Z}/4\mathbb{Z}] \to \mathbb{Z}$.

Definition. *For any group ring, the map $R[G] \to R$ sending $g \mapsto 1$ (or -1) for all $g \in G$ is called the augmentation.*

Using the previous example, we can construct a map

$$\mathbb{Z}[\mathbb{Z}/4\mathbb{Z}] \to \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}[i]$$

given by

$$x \mapsto (1, -1, i)$$

For instance $1 + x^2 \mapsto (2, 2, 0)$. Both sides are isomorphic to $\mathbb{Z}^4$ but yet this map is not an isomorphism (we can't get $(1, 0, 0)$). However if we consider a complex group ring $\mathbb{C}[G]$ instead then we can get an isomorphism

$$\mathbb{C}[\mathbb{Z}/4\mathbb{Z}] = \mathbb{C} \oplus \mathbb{C} \oplus \mathbb{C} \oplus \mathbb{C}$$

given by the map

$$x \mapsto (1, -1, i, -i)$$

For this to be an isomorphism there must exist a polynomial f such that

$$f(1) = z_1 \quad f(-1) = z_{-1} \quad f(i) = z_i \quad f(-i) = z_{-i}$$

for any 4-tuple $(z_1, z_{-1}, z_i, z_{-i})$. Through polynomial interpolation we can construct a degree 3 polynomial satisfying these constraints so

$$\mathbb{C}[\mathbb{Z}/4\mathbb{Z}] \cong \mathbb{C} \oplus \mathbb{C} \oplus \mathbb{C} \oplus \mathbb{C}$$

In general we can break up $\mathbb{C}[G]$ for any finite group G, such rings are special.

Definition. *A ring R is semisimple if any short exact sequence of left R-modules splits, that is if we have an exact sequence*

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

then there exists a map (called a section) $s : C \to B$ such that $g \circ s : C \to C$ is the identity map. It thus follows that

$$B \cong A \oplus C \text{ or more precisely } B = f(A) \oplus s(C)$$

In particular if B has a submodule A, then we have the exact sequence

$$0 \longrightarrow A \xrightarrow{\iota} B \xrightarrow{\pi} B/A \longrightarrow 0$$

and so A is a direct summand of B, that is $B = A \oplus A'$ for some other submodule A'.

Thus our discussion so far has shown that $\mathbb{C}[G]$ is a semisimple ring. The augmentation $\alpha : \mathbb{C}[G] \to \mathbb{C}$ given by $\alpha(g) = 1$ for all $g \in G$ makes $\mathbb{C}$ into a $\mathbb{C}[G]$-module under the action $gz = \alpha(g)z$. In other words this is the trivial representation of G.

Suppose V is also a representation of G, considered as a $\mathbb{C}[G]$-module. The tensor product $\mathbb{C} \otimes V$ is spanned by

$$z \otimes v = z(1 \otimes v)$$

So as a $\mathbb{C}$-vector space, the tensor product is spanned by $1 \otimes v$ which can be thought of as vectors in V. For some $g \in G$

$$1 \otimes gv = \alpha(g) \otimes v = 1 \otimes v$$

These are subject to the relation $gv = v \to (g-1)v = 0$ since all g act trivially. This is known as a coinvariant quotient.

Another way to view this is if $I = \ker \alpha = (g-1)$ for all $g \in G$. Then $\mathbb{C} = \mathbb{C}[G]/I$ and thus the tensor is (by an earlier discussion)

$$\mathbb{C} \otimes v = \mathbb{C}[G]/I \otimes V = V/IV = V/(g-1)v$$

Note that we are quotienting by the set $\{(g-1)v \mid g \in G, v \in V\}$.

To see what this is precisely consider the linear map $q : V \to V_G$ where V_G is the coinvariant quotient. We can think of V_g as a representation of G in which G acts trivially. $q(gv) = q(v)$ since $q((g-1)v) = 0$. Thus

$$q(gv) = gq(v) = q(v)$$

and this is a surjective map of representations of G from V to some number of copies of the trivial representation.

If V is irreducible and nontrivial, then $V_g = 0$ by Schur's Lemma. If V is trivial then $(g-1)v = 0$ to begin with and so $V_G = V$. For a general V

$$V \cong \bigoplus_{V_i \text{ irrep}} V_i^{\alpha_i}$$

In this case V_G is the quotient of V by all its nontrivial irreducible constituents (i.e. it zeroes them out), this is known as the trivial-isotypical part of V and has dimension

$$\dim V_G = \langle V, triv \rangle$$

Returning to our discussion of groups rings and now we can consider cases where R is noncommutative as well, the following discussion applies for all rings.

Definition. *The Jacobson radical is the intersection of all maximal ideals of R and is denoted $J(R)$*

Example. $J(\mathbb{Z}) = \bigcap p\mathbb{Z} = 0$

Example. For a local ring like $\mathbb{C}[[t]]$ with only one maximal ideal, $J(\mathbb{C}[[t]]) = (t)$

Lemma 7.1 (Nakayama)**.** Let M be a finitely generated left R-module such that $JM = M$ (J denotes the Jacobson radical $J(R)$), then $M = 0$.

Proof. Suppose M is generated by $m_1, \dots, m_n$ with n as small as possible. By assumption $M = JM$, in particular $m_1 \in JM$ and so

$$m_1 = j_1 m_1 + \cdots + j_n m_n \to (1 - j_1)m_1 = j_2 m_2 + \cdots + j_n m_n \quad j_i \in J$$

But for every maximal ideal $I \subset R$, $j_1 \in I$ due to the properties of the Jacobson radical. Thus $1 - j_1 \notin I$ or else $1 \in I$ and it is no longer maximal. This implies that $(1 - j_1)$ is not contained in any maximal ideal and so must be the unit ideal which means $1 - j_1$ is invertible, thus

$$m_1 = (1 - j_1)^{-1}(j_2 m_2 + \cdots + j_n m_n)$$

this is a contradiction because $m_1, \dots, m_n$ are linearly independent so $M = 0$. □

We can compare this result with the fact that in a local ring, the unique maximal ideal is the set of noninvertible elements. The Jacobson radical J is the analog for nonlocal rings.

Corollary 7.2. Suppose $m_1, \dots, m_k \in M$ are elements such that their images $\overline{m_1}, \dots, \overline{m_k} \in M/JM$ generate M/JM, then $m_1, \dots, m_k$ generate M.

Proof. Let $N = Rm_1 + \cdots + Rm_k$ be the generated submodule. By assumption, the sequence

$$N \to M \to M/JM$$

is surjective, in other words $N + JM = M$. Projecting onto M/N and using the lemma gives

$$0 + J(M/N) = M/N \to M/N = 0 \to M = N$$

so the the elements generate M. □

Example. Consider $R = \mathbb{C}[[t]]$ and $M = \mathbb{C}[[t]]^k$ and suppose we are given some $f_1, \dots, f_k \in M$ and we would like to know if they generate M. Let $\overline{f_i}$ be the constant terms of f_i i.e. the image under $M/(t)M = (\mathbb{C}[[t]]/(t))^k = \mathbb{C}^k$. By the corollary it suffices to show that $\overline{f_1}, \dots, \overline{f_k}$ generate $\mathbb{C}^k$ which is way simpler, we just need to show they are linearly independent.

The requirement that M be finitely generated is very important.

Example. Let $M = \mathbb{C}((t)) = \sum_{i=-m}^{\infty} a_i x^i$ be the field of Laurent series. This is a $\mathbb{C}[[t]]$-module which has a Jacobson radical of $J(\mathbb{C}[[t]]) = (t)$. $tM = M$ but yet $M \neq 0$ which is because M is not finitely generated and so Nakayama's lemma does not apply.

We can use Jacobson radicals to give another characterization of semisimple rings.

Definition. *A ring is Artinian if it satisfies the descending chain condition. That is the chain*

$$I_1 \supset I_2 \supset I_3 \supset \cdots$$

must terminate.

Example. $\mathbb{Z}$ is not Artinian because $(2) \supset (4) \supset (8) \supset \cdots$

Example. A frequently encountered class of Artinian rings is the case where R contains a field k and $\dim_k R < \infty$. In this case every ideal of R is a subspace of R as a k-vector space and we can associate

$$I_1 \supset I_2 \supset \cdots \Longleftrightarrow V_1 \supset V_2 \supset \cdots$$

Since $\dim V_i$ is strictly decreasing this chain must terminate. We can apply this to

$$M_n(k) : \dim_k M_n(k) = n^2 \qquad k[G] : \dim_k k[G] = |G|$$

Theorem 7.3. A ring is semisimple if and only if it is Artinian with a trivial Jacobson radical.

From this we see that $k[G]$ and $M_n(k)$ are semisimple when k is a field. The latter fact follows from the correspondence between $M_n(k)$-modules and k-vector spaces. It's interesting to note that Nakayama's lemma is a case for when the Jacobson radical is "too large." In the case where it is very small (trivial in fact), the ring is semisimple. The structure of semisimple rings is given by a powerful theorem.

Theorem 7.4 (Artin-Wedderburn)**.** Let R be a semisimple ring. Then R is a direct sum of algebras of the form $M_n(D)$ where D is a division ring. If R contains a field k (i.e. is a k-algebra), then each D is also a k-algebra.

Corollary 7.5. Suppose R is a semisimple ring which is also a k-algebra with K algebraically closed and $\dim_k R < \infty$, then

$$R = \bigoplus_i M_{n_i}(k)$$

Proof. By the Artin-Wedderburn Theorem

$$R = \bigoplus_i M_{n_i}(D_i)$$

so we just need to show that $D_i = k$ for all i. Let $a \in D_i \setminus k$ and so $1, a, a^2, \ldots$ are linearly related since $\dim_k D < \infty$. In other words there is some polynomial f such that $f(a) = 0$, let f be of minimal degree. f must have some root $\lambda \in k$, that is

$$f(x) = (x - \lambda)g(x) \to f(a) = (a - \lambda)g(a) = 0$$

But $(a - \lambda) \neq 0$ because $a \notin k$ and $g(a) \neq 0$ because f is the least degree in which $f(a) = 0$. Thus we arrive at a contradiction and so $D_i \subseteq k$, but D_i is a k-algebra and so contains k, therefore $D_i = k$ for all i. □

Example. We've already shown that $\mathbb{C}[G]$ is semisimple. Since $\mathbb{C}$ is algebraically closed we get

$$\mathbb{C}[G] \cong \bigoplus_i M_{n_i}(\mathbb{C})$$

We would like to know what the summands actually are. Recall that representation theory is the study of homomorphisms

$$\rho : G \to GL_n(\mathbb{C})$$

which induces a ring homomorphism

$$\mathbb{C}[G] \to M_n(\mathbb{C})$$

In particular if V is an irreducible representation then

$$\mathbb{C}[G] \to End(V) \cong M_n(\mathbb{C}) \qquad n = \dim V$$

Let $V_1, \ldots, V_k$ be the irreducible representations of G, then

$$\mathbb{C}[G] \to \bigoplus_{i=1}^{k} End(V_i) \cong \bigoplus_{i=1}^{k} M_{n_i}(\mathbb{C}) \qquad n_i = \dim V_i$$

This map is the isomorphism generated by the Artin-Wedderburn Theorem

$$\dim \oplus M_{n_i}(\mathbb{C}) = \sum \dim M_{n_i}(\mathbb{C}) = \sum n_i^2 = |G|$$

which was an identity previously derived through character theory.

Finally let's consider the center $Z(\mathbb{C}[G])$. On one hand this is spanned by $\sum_{g \in c} g$ where c is a conjugacy class. This implies

$$\dim Z(\mathbb{C}[G]) = \text{number of conjugacy classes}$$

On the other hand we have

$$Z(\bigoplus_{i=1}^{k} M_{n_i}(\mathbb{C})) = \bigoplus_{i=1}^{k} \mathbb{C} \cdot I_{n_i}$$

where I is the identity matrix and this has dimension k, which is the number of irreducible representations. Thus we've derived another result, the number of irreducible representations is equal to the number of conjugacy classes.

7.2 Rings of Fractions

One of the most important tools in commutative algebra is the formation of fractions and, by extension, the process of localization. For a familiar example let's think closely at how we can form $\mathbb{Q}$ from $\mathbb{Z}$. First note that 1 should always be invertible in a ring, furthermore if x, y are invertible then so is their product. Immediately we can start generalizing the formation of fractions.

Definition. *Let A be a commutative ring, a subset $S \subseteq A$ is a multiplicative set if*

1. $1 \in S$
2. $x, y \in S \to xy \in S$

Clearly the set of nonzero integers is a multiplicative set (hence why we don't divide by zero in a fraction), let's consider another example.

Example. If $x \in A$, then we have the multiplicative set

$$S = \{1, x, x^2, x^3, \ldots\}$$

If $\mathfrak{p} \subseteq A$ is prime, then we have another multiplicative set

$$S = A \setminus \mathfrak{p}$$

Now that we have the set of possible denominators (the multiplicative set), the set of rational numbers (fractions of $\mathbb{Z}$) can be defined as

$$\mathbb{Q} = \{(m, n) \in \mathbb{Z}^2 \mid n \in S\}$$

where we take the multiplicative set to be $S = \mathbb{Z} \setminus \{0\}$. However note that in this definition $4/6$ and $2/3$ are distinct fractions when they should be equal. To rectify this we must define an equivalence relation

$$(m, n) \sim (m', n') \to mn' = m'n$$

Modding by this relation will give the full set of fractions

$$\mathbb{Q} = \left\{(m, n) = \frac{m}{n} \in \mathbb{Z}^2 \mid n \in S\right\} / \sim$$

with the ring structure

$$\frac{m}{n} + \frac{m'}{n'} = \frac{mn' + m'n}{nn'} \qquad \frac{m}{n} \cdot \frac{m'}{n'} = \frac{mm'}{nn'}$$

which we expect to get from fractions. Note that this also defines the map

$$\mathbb{Z} \to \mathbb{Q} \qquad m \mapsto \frac{m}{1}$$

furthermore if $m \in S$, then m is invertible ($m/1 \to 1/m$).

Definition. *Let A be a commutative ring and $S \subseteq A$ a multiplicative set, then the ring of fractions is*

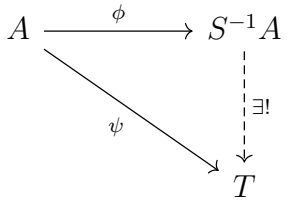
$$S^{-1}A = \left\{\frac{m}{n} \mid m \in A, n \in S\right\} / \sim$$

with the equivalence relation

$$\frac{m}{n} \sim \frac{m'}{n'} \to \exists t \in S \text{ s.t. } (mn' - m'n)t = 0$$

The t has to be included in the equivalence relation because A may have zero divisors. Note that $S^{-1}A$ satisfies the universal property.

1. There exists a map $\phi : A \to S^{-1}A$ given by $\phi(x) = x/1$
2. For all $s \in S$, $\phi(s)$ is invertible in $S^{-1}A$
3. Thus $S^{-1}A$ is universal as



where ψ is a ring map such that $\psi(S)$ is a subset of the units of T.

The multiplicative sets we introduced earlier form special rings of fractions, thus we give them their own notation.

- For $x \in A$ and the multiplicative set $S = \{1, x, x^2, \ldots\}$ we will denote

$$A_x = S^{-1}A$$

- For $\mathfrak{p} \subseteq A$ prime and $S = A \setminus \mathfrak{p}$ we denote

$$A_{\mathfrak{p}} = S^{-1}A$$

Example. To illustrate the difference between the two rings above, let p be prime

$$\mathbb{Z}_p = \left\{ \frac{m}{p^k} \mid m, k \in \mathbb{Z} \right\} \subseteq \mathbb{Q}$$
$$\mathbb{Z}_{(p)} = \left\{ \frac{m}{n} \mid m, n \in \mathbb{Z}, p \nmid n \right\} \subseteq \mathbb{Q}$$

Observe that $A_{\mathfrak{p}}$ is a local ring with unique maximal ideal

$$\mathfrak{p}A_{\mathfrak{p}} = \left\{ \frac{m}{n} \mid m \in \mathfrak{p}, n \notin \mathfrak{p} \right\}$$

thus $A_{\mathfrak{p}}$ is called the localization of A at $\mathfrak{p}$.

So now we can use multiplicative sets to form rings of fractions, but can we do one better? Why stop at rings when we can make modules?

Definition. *Let A, S be as defined before and M be an A-module, then*

$$S^{-1}M = \left\{ \frac{m}{s} \mid, \in M, s \in S \right\} / \sim$$

with the same equivalence relation as before. This turns an A-module M into a $S^{-1}A$-module with obvious definitions of addition and scalar multiplication.

Theorem 7.6. The operation S^{-1} is exact, furthermore

$$S^{-1}M \cong M \otimes_A S^{-1}A$$

This shows that $S^{-1}A$ is a flat A-module

Proof. Suppose we have the exact sequence

$$M' \xrightarrow{f} M \xrightarrow{g} M''$$

and form another sequence

$$S^{-1}M' \xrightarrow{S^{-1}f} S^{-1}M \xrightarrow{S^{-1}g} S^{-1}M''$$

where we've defined the map

$$(S^{-1}f)\left(\frac{m}{s}\right) = \frac{f(m)}{s}$$

Note that since $g \circ f = 0$ then $S^{-1}g \circ S^{-1}f = 0$ as well. This proves half of the equality we want, namely

$$\text{im } S^{-1}f \subseteq \ker S^{-1}g$$

For the other side, let $m/s \in \ker S^{-1}g$ so that

$$(S^{-1}g)\left(\frac{m}{s}\right) = 0 \to \frac{g(m)}{s} = 0 \in S^{-1}M''$$

This is equivalent to saying that $g(m)/s \sim 0/1$ and thus there must exist some $t \in S$ such that $tg(m) = g(tm) = 0 \to tm \in \ker g$. But the original sequence was exact, so $\ker g = \text{im } f$, thus we can find some $m' \in M'$ such that $f(m') = tm$. In other words

$$(S^{-1}f) = \left(\frac{m'}{ts}\right) = \frac{f(m')}{ts} = \frac{tm}{ts} = \frac{m}{s}$$

which demonstrates $\ker S^{-1}g \subseteq \text{im } S^{-1}f$, proving the desired equality and showing that S^{-1} is exact.

To prove the isomorphism we can use a bilinear map

$$f : M \times S^{-1}A \to S^{-1}M \qquad (m, a/s) \mapsto ma/s$$

which can be extended to a map

$$f : M \otimes S^{-1}A \to S^{-1}M \qquad m \otimes a/s \mapsto ma/s$$

We can construct another map

$$g : S^{-1}M \to M \otimes S^{-1}A \qquad m/s \mapsto m \otimes 1/s$$

which can be shown to be the inverse to f, thus establishing the isomorphism. □

Corollary 7.7. If N, P are submodules of M, then

1. $S^{-1}(N+P) = S^{-1}N + S^{-1}P$
2. $S^{-1}(N \cap P) = S^{-1}N \cap S^{-1}P$
3. $S^{-1}(M/N) = S^{-1}M/S^{-1}N$

Thus formation of fractions commutes with finite sums, intersections, and quotients.

Proof. 1) and 2) are easy to verify from the definitions and by applying the equivalence relation. For 3) apply S^{-1} to the exact sequence

$$0 \longrightarrow N \longrightarrow M \longrightarrow M/N \longrightarrow 0$$

□

We briefly discussed the idea of localizations earlier; this is an extremely powerful concept, especially in the field of algebraic geometry. As such, properties preserved by the process of localization are in a sense "special."

Definition. *A property P of a ring A (or A-module M) is said to be a local property if A has P if and only if $A_\mathfrak{p}$ has P for all prime ideals $\mathfrak{p}$.*

In other words, A has some property only if all of its localizations have that same property. We'll explore some simple local properties in the next few propositions.

Proposition 7.8. Let M be an A-module, then the following are equivalent

1. $M = 0$
2. $M_\mathfrak{p} = 0$
3. $M_\mathfrak{m} = 0$

Proof. $1 \to 2 \to 3$ is obvious, suppose 3 holds and $M \neq 0$ for the purpose of contradiction. Let $x \in M$ be nonzero and $\mathfrak{a} = Ann(x)$ which is a nontrivial ideal. $\mathfrak{a}$ is contained in some maximal ideal $\mathfrak{m}$ so consider $x/1 \in M_\mathfrak{m} = 0$, thus x is annihilated by some elements of $A \setminus \mathfrak{m}$ according to the equivalence relation. This is a contradiction as $Ann(x) \subseteq \mathfrak{m}$. □

Proposition 7.9. Let $\phi : M \to N$ be an A-module homomorphism, then the following are equivalent

1. ϕ is injective
2. $\phi_\mathfrak{p} : M_\mathfrak{p} \to N_\mathfrak{p}$ is injective
3. $\phi_\mathfrak{m} : M_\mathfrak{m} \to N_\mathfrak{m}$ is injective

Proof. $1 \to 2 \to 3$ is obvious. Suppose that 3 holds and let $M' = \ker \phi$. We can form the exact sequence

$$0 \longrightarrow M' \longrightarrow M \xrightarrow{\phi} N$$

Since S^{-1} is exact, we can also form the exact sequence

$$0 \longrightarrow M'_{\mathfrak{m}} \longrightarrow M_{\mathfrak{m}} \xrightarrow{\phi_{\mathfrak{m}}} N_{\mathfrak{m}}$$

Thus $M'_{\mathfrak{m}} \cong \ker(\phi_{\mathfrak{m}} = 0$ which means that $M' = 0$ by the previous proposition and thus ϕ is injective. □

The same proof can be used for surjectivity.

Proposition 7.10. Let M be an A-module, then the following are equivalent

1. M is flat
2. $M_{\mathfrak{p}}$ is flat
3. $M_{\mathfrak{m}}$ is flat

Proof. $1 \to 2$ follows from theorem 7.6, $2 \to 3$ is obvious. Suppose 3 holds, let $N \to P$ be an injective A-module homomorphism and $\mathfrak{m}$ a maximal ideal.

$$\begin{aligned} N \to P \text{ injective} &\Rightarrow N_{\mathfrak{m}} \to P_{\mathfrak{m}} \text{ injective (by 12.9)} \\ &\Rightarrow N_{\mathfrak{m}} \otimes_{A_{\mathfrak{m}}} M_{\mathfrak{m}} \to P_{\mathfrak{m}} \otimes_{A_{\mathfrak{m}}} M_{\mathfrak{m}} \text{ injective } (M_{\mathfrak{m}} \text{ is flat}) \\ &\Rightarrow (N \otimes_A M)_{\mathfrak{m}} \to (P \otimes_A M)_{\mathfrak{m}} \text{ injective (by 12.6)} \\ &\Rightarrow N \otimes_A M \to P \otimes_A M \text{ injective (by 12.9)} \end{aligned}$$

Thus M is flat. □

To summarize, the following are all local properties

1. is trivial $(= 0)$
2. is injective/surjective
3. is flat

Although we haven't discussed much about ideals in rings of fractions, for most purposes it is enough to remember that we have a natural map

$$f : A \to S^{-1}A$$

and so we can consider extended/contracted ideals. In fact all ideals of $S^{-1}A$ will be an extended ideal of some $\mathfrak{a} \subseteq A$ and thus of the form $S^{-1}\mathfrak{a}$.

7.3 Primary Decompositions

Similar to how numbers can be factored into their prime factors, the process of decomposing an ideal into primary ideals was a pillar of algebraic geometry. While no longer a central part of the theory which now focuses on localization, this is still a useful topic to be familiar with.

A prime ideal is in a sense a generalization of prime numbers, the primary ideal is the generalization of a prime power.

Definition. *An proper ideal* $\mathfrak{q}$ *is primary if*

$$xy \in \mathfrak{q} \to x \in \mathfrak{q} \text{ or } y^n \in \mathfrak{q} \text{ for some } n > 0$$

or equivalently every zero divisor of $A/\mathfrak{q}$ *is nilpotent.*

Compare this to the case of prime ideals, if $\mathfrak{p}$ is prime then $A/\mathfrak{p}$ is an integral domain and so the zero divisors are actually zero rather than just nilpotent. From this definition we can see that every prime ideal and contractions of prime ideals are primary.

Proposition 7.11. If $\mathfrak{q}$ is primary, then $r(\mathfrak{q})$ is the smallest prime ideal which contains $\mathfrak{q}$.

Proof. Since $r(\mathfrak{q})$ can be thought of as the intersection of all prime ideals which contain $\mathfrak{q}$, we just need to show that $\mathfrak{p} = r(\mathfrak{q})$ is prime to complete the proof. Let $xy \in r(\mathfrak{q})$ which implies $(xy)^n \in \mathfrak{q}$ for some $n > 0$. In other words, either $x^n \in \mathfrak{q}$ or $y^{mn} \in \mathfrak{q}$ for some $m > 0$, thus either $x \in r(\mathfrak{q})$ or $y \in r(\mathfrak{q})$. □

Definition. *We say that* $\mathfrak{q}$ *is* $\mathfrak{p}$*-primary if* $\mathfrak{p} = r(\mathfrak{q})$.

We can think of a $\mathfrak{p}$-primary ideal as be a prime power of $\mathfrak{p}$, after all the primary ideals are generalizations of prime powers. Since the standard prototypes of commutative rings are $\mathbb{Z}$ and $k[x_1, \ldots, x_n]$ for some field k, we will start our examples there.

Example. The primary ideals of $\mathbb{Z}$ are (0) and (p^n) where p is prime, which we can immediately check because their radicals are prime.

Example. Let $A = k[x, y]$ and consider $\mathfrak{q} = (x, y^2)$. The quotient is

$$A/\mathfrak{q} \cong k[y]/(y^2)$$

in which the zero divisors are of the form $f(x)y^n$ and thus are all nilpotent. Therefore $\mathfrak{q}$ is a primary ideal, note that we have

$$r(\mathfrak{q}) = (x, y) \qquad \mathfrak{p}^2 = (x^2, y^2) \subset \mathfrak{q} \subset p$$

From these two examples we see that while the primary ideals are generalizations of prime (number) powers, they are by themselves not necessarily a prime (ideal) power. Conversely a prime power $\mathfrak{p}^n$ is not necessarily primary even though its radical is prime ($= \mathfrak{p}$). However all is not lost as we have the following result.

Proposition 7.12. If $r(\mathfrak{a})$ is maximal, then $\mathfrak{a}$ is primary. In particular, the powers of a maximal ideal $\mathfrak{m}$ are $\mathfrak{m}$-primary.

Proof. Let $\mathfrak{m} = r(\mathfrak{a})$, the image of $\mathfrak{m}$ in $A/\mathfrak{a}$ is its nilradical. But since $\mathfrak{m}$ is already maximal, $A/\mathfrak{a}$ must have only one prime ideal $(\overline{\mathfrak{m}})$. Every elements of $A/\mathfrak{a}$ is thus either a unit or nilpotent, so all the zero divisors are nilpotent. □

Before we begin the main topic of this section, we have a couple of lemmas to make ensure that everything we say makes sense.

Lemma 7.13. The intersection of $\mathfrak{p}$-primary ideals is $\mathfrak{p}$-primary

Proof. Follows directly from properties of the ideal radical. □

Lemma 7.14. Let $\mathfrak{q}$ be $\mathfrak{p}$-primary and $x \in A$, then

- If $x \in \mathfrak{q}$, then $(\mathfrak{q} : x) = (1)$
- If $x \notin \mathfrak{q}$, then $(\mathfrak{q} : x)$ is $\mathfrak{p}$-primary and thus $r(\mathfrak{q} : x) = \mathfrak{p}$
- If $x \notin \mathfrak{p}$, then $(\mathfrak{q} : x) = \mathfrak{q}$

Proof. 1) and 3) are clear from definitions. For 2), if $y \in (\mathfrak{q} : x)$ then $xy \in \mathfrak{q} \to y^n \in \mathfrak{q} \to y \in \mathfrak{p}$. Thus we get

$$\mathfrak{q} \subseteq (\mathfrak{q} : x) \subseteq \mathfrak{p} \to r(\mathfrak{q} : x) = \mathfrak{p}$$

If $yz \in (\mathfrak{q} : x)$ with $y \notin \mathfrak{p}$, then $xyz \in \mathfrak{p}$ and thus $xz \in \mathfrak{q} \to z \in (\mathfrak{q} : x)$. □

Definition. *A primary decomposition of an ideal $\mathfrak{a}$ is an expression of $\mathfrak{a}$ as a finite intersection of primary ideals*

$$\mathfrak{a} = \bigcap_{i=1}^{n} \mathfrak{q}_i$$

In general these decompositions need not exist, however that would certainly be boring so we focus on ideals which have a primary decomposition for now.

Definition. *If we have a decomposition where all the radicals $r(\mathfrak{q}_i)$ are distinct and furthermore no component contains the intersection of all other components*

$$\mathfrak{q}_i \not\supseteq \bigcap_{i \neq j} \mathfrak{q}_j$$

then we say the decomposition is minimal (or irredundant, or reduced).

By using lemma 7.13 we can combine ideals by intersecting them until we get all distinct radicals and we can eliminate redundant terms until we get a minimal decompositions. In other words, a primary decomposition can always be reduced to a minimal one. If an ideal has a primary decomposition we will say that it is decomposable.

Now let's talk uniqueness

Theorem 7.15 (1st Uniqueness Theorem)**.** Let $\mathfrak{a}$ be a decomposable ideal with a minimal primary decomposition

$$\mathfrak{a} = \bigcap_{i=1}^{n} \mathfrak{q}_i$$

Let $\mathfrak{p}_i = r(\mathfrak{q}_i)$, then these are just the prime ideals which occur in the set of ideals

$$\{r(\mathfrak{a} : x) \mid x \in A\}$$

and are thus independent of the particular decomposition of $\mathfrak{a}$.

Proof. For any $x \in A$, note that

$$(\mathfrak{a} : x) = (\cap \mathfrak{q}_i : x) = \cap (\mathfrak{q}_i : x)$$

$$\therefore r(\mathfrak{a} : x) = \bigcap_{i=1}^{n} r(\mathfrak{q}_i : x) = \bigcap_{x \notin \mathfrak{q}_j} \mathfrak{p}_j$$

If this is prime, then $r(\mathfrak{a} : x) = \mathfrak{p}_j$ for some j and thus every prime of this form is one of the $\mathfrak{p}_j$. Conversely, for each i there exists some $x_i \notin \mathfrak{q}_i$ which is in $\cap_{j \neq i} \mathfrak{q}_j$ since the decomposition is minimal; thus we get $r(\mathfrak{a} : x_i) = \mathfrak{p}_i$. □

The above theorem implies that while the decomposition may not be unique, the underlying set of prime ideals (the radicals) will remain the same in a minimal decomposition. If we consider $A/\mathfrak{a}$ as an A-module, the theorem is equivalent to saying that the $\mathfrak{p}_i$ are just the prime ideals which occur as radicals of annihilators of elements of $A/\mathfrak{a}$.

Example. Consider $\mathfrak{a} = (x^2, xy) \subseteq k[x, y]$. This can be decomposed as

$$\mathfrak{a} = \mathfrak{p}_1 \cap \mathfrak{p}_2^2 = (x) \cap (x, y)^2$$

Note that in this example we have $\mathfrak{p}_1 \subset \mathfrak{p}_2$ with

$$r(\mathfrak{a}) = \mathfrak{p}_1 \cap \mathfrak{p}_2 = \mathfrak{p}_1$$

which is prime but $\mathfrak{a}$ is not primary.

Definition. *In the theorem, the prime ideals $\mathfrak{p}_i$ are said to belong to (or associated with) $\mathfrak{a}$. An ideal is prime if and only if it has only one associated prime ideal.*

The minimal elements of the set $\{\mathfrak{p}_1, \ldots, \mathfrak{p}_n\}$ are called the minimal (or isolated) prime ideals belonging to $\mathfrak{a}$. The other ideals are called embedded.[6]

Proposition 7.16. Let $\mathfrak{a}$ be decomposable, then any prime ideal $\mathfrak{p} \supseteq \mathfrak{a}$ contains a minimal prime ideal belonging to $\mathfrak{a}$. Thus the minimal prime ideals of $\mathfrak{a}$ are just the minimal elements in the set of all prime ideals which contain $\mathfrak{a}$.

Proof.

$$\mathfrak{p} \supseteq \mathfrak{a} = \bigcap_{i=1}^{n} \mathfrak{q}_i \to \mathfrak{p} = r(\mathfrak{p}) \supseteq \cap r(\mathfrak{q}_i) = \cap \mathfrak{p}_i$$

Thus $\mathfrak{p} \supseteq \mathfrak{p}_i$ for some i and so contains a minimal prime ideal of $\mathfrak{a}$. □

[6] These names come from algebraic geometry in the study of varieties

7.4 Integral Dependence

Sometimes we encounter in our studies a "bad" ring which is "unpleasant" to work with. In this section we will give a concrete definition of what makes rings "bad" and methods in which we can turn them into "good" rings.

Definition. *Let $A \subseteq B$ be rings, an elements $x \in B$ is integral if there exists coefficients $a_1, \ldots, a_n \in A$ such that*

$$x^n + a_1 x^{n-1} + \cdots + a_n = 0$$

(with computation carried out in B).

Note that elements of A are trivially integral over A.

Example. Consider $\mathbb{Z} \subseteq \mathbb{Q}$ and suppose $r/s \in \mathbb{Q}$ is integral over $\mathbb{Z}$. Furthermore suppose that this fraction is reduced, that is $gcd(r, s) = 1$. Then

$$\left(\frac{r}{s}\right)^n + a_1 \left(\frac{r}{s}\right)^{n-1} + \cdots + a_n = 0$$
$$r^n + a_n s r^{n-1} + \cdots + a_n s^n = 0$$

which implies that s divides r^n and thus $s = \pm 1 \to r/s \in \mathbb{Z}$. Thus the only integral elements over $\mathbb{Z}$ are $\mathbb{Z}$ itself. Rings with this property have a special name which we will see later.

Definition. *An A-module M is faithful if $a \cdot M = 0$ implies $a = 0$.*

Proposition 7.17. The following are equivalent

1. $x \in B$ is integral over A
2. $A[x]$ is a finitely generated A-module
3. $A[x]$ is contained in a subring $C \subseteq B$ such that C is a finitely generated A-module
4. There exists a faithful $A[x]$-module M which is finitely generated as an A-module

Proof. $1 \to 2$: If x is integral, then we have for all $r \geq 0$

$$x^{n+r} = (a_1 x^{n+r-1} + \cdots + a_n x^r$$

so that $A[x]$ is finitely generated by $\{1, x, \ldots, x^{n-1}\}$.

$2 \to 3$: Take $C = A[x]$.

$3 \to 4$: Tkae $M = C$ which is faithful because if $yC = 0$ then $y \cdot 1 = 0$.

$4 \to 1$: Suppose M is a finitely generated (as an A-module) faithful $A[x]$-module and consider the endomorphism

$$\phi : M \to M \qquad m \mapsto mx$$

There must exist $a_1, \ldots, a_n \in A$ such that

$$\phi^n + a_1\phi^{n-1} \cdots + a_n = 0$$

as endophormisms (property of finitely generated modules). This essentially implies that

$$x^n + a_1 x^{n-1} + \cdots a_n = 0$$

which means x is integral over A. □

Corollary 7.18. If $x_1, \ldots, x_n$ are integral over A, then $A[x_1, \ldots, x_n]$ is a finitely generated A-module.

Proof. This can be shown using induction on n and repeatedly applying the proposition. □

Corollary 7.19. The set

$$C = \{x \in B \mid x \text{ integral over } A\}$$

is a subring of B containing A.

Proof. If $x, y \in C$ then $A[x, y]$ is finitely generated. But $x + y, xy \in A[x, y]$ so

$$A[x, y], A[xy] \subseteq A[x, y]$$

so both are finitely generated and thus integral. □

Definition. *The ring C in the corollary above is called the integral closure of A in B. If the only elements integral over A are the trivial ones ($C = A$), then we say A is integrally closed in B. If all elements are integral over A ($C = B$), then we say B is integral over A.*

Example. From the earlier example we've shown that $\mathbb{Z}$ is integrally closed in $\mathbb{Q}$.

Some remarks about integral module and algebras:

Definition. *B is an A-algebra of finite type if it is finitely generated. This is equivalent to saying that the smallest subring of B containing A is B itself.*

If $A \subseteq B$ are rings and B is a finitely generated A-module, then B is finite over A.

In the case of an A-algebra we have a ring homomorphism $f : A \to B$, if B is integral over it's subring $f(A)$ then we say that f is integral and B is an integral A-algebra. Using the terminology we introduced, our results prove

$$\text{finite type} \ + \ \text{integral} \ = \ \text{finite}$$

Now we will examine some properties of integral dependence. First it can be shown that it is transitive.

Corollary 7.20. If $A \subseteq B \subseteq C$ with C integral over B and B integral over A, then C is integral over A.

Proof. Let $x \in C$ and

$$x^n + b_1 x^{n-1} + \cdots + b_n = 0$$

The ring $B = A[b_1, \ldots, b_n]$ is a finitely generated A-module and $B'[x]$ is a finitely generated B'-module (since x is integral over B'). Thus it is also a finitely generated A-module and so x is integral over A. □

Corollary 7.21. Let $A \subseteq B$ and C be the integral closure of A in B, then C is integrally closed in B.

Proof. If $x \in B$ is integral over C, then x is also integral over A (since $A \subseteq C \subseteq B$) and thus $x \in C$. □

This corollary proves that the integral closure is indeed integrally closed. Integral dependence is preserved under certain operations which we will prove in the following proposition.

Proposition 7.22. Let $A \subseteq B$ with B integral over A, then

1. If $\mathfrak{a} = \mathfrak{b}^c = A \cap \mathfrak{b}$, then $B/\mathfrak{b}$ is integral over $A/\mathfrak{a}$
2. $S^{-1}B$ is integral over $S^{-1}A$

Proof. 1) For some $x \in B$ suppose

$$x^n + \cdots + a_n = 0$$

We can just reduce everything mod $\mathfrak{b}$ do get our desired result.

2) If $x/s \in S^{-1}B$, then

$$\left(\frac{x}{s}\right)^n + \left(\frac{a_1}{s}\right)\left(\frac{x}{s}\right)^{n-1} + \cdots + \left(\frac{a_n}{s^n}\right) = 0$$

so that x/s is integral over $S^{-1}A$. □

The rest of this section will be devoted to discussing chains of ideals in integral closures. The major results in this area are the Coden-Seidenberg theorems, commonly called the going-up and going-down theorems, though we will mostly focus on the going-up theorem here.

Lemma 7.23. Let $A \subseteq B$ be integral domains with B integral over A. Then B is a field if and only if A is also a field.

Proof. Suppose A is a field and let $b \in B$ be nonzero. Since B is integral over A we have the equation

$$b^n + a_1 b^{n-1} + \cdots + a_n = 0$$

Note that $a_n \neq 0$ because otherwise

$$b(b^{n-1} + a_1 b^{n-2} + \cdots + a_1) = 0$$

where $b \neq 0$ would create a smaller degree integral dependence. THus we assume $a_n \neq 0$ and write

$$b^n + \cdots + a_1 b = -a_n$$
$$b \cdots (-a_n^{-1}(b^{n-1} + \cdots + a_{n-1})) = 1$$

and thus b is invertible.

Conversely, let B be a field and $a \in A$ be nonzero. Since $A \subseteq B$, a is invertible in B and since B is integral over A we can write

$$(a^{-1})^n + a_1(a^{-1})^{n-1} + \cdots + a_n = 0$$

If we multiply by a^{n-1} we get

$$a^{-1} + a_1 + \cdots + a_n a^{n-1} = 0$$
$$a^{-1} = -(a_1 + \cdots + a_n a^{n-1}) \in A$$

and so a is invertible in A. □

This is just one example of a property which is passed through integral dependence. We will look at a few more in the following results.

Corollary 7.24. Let $A \subseteq B$ be rings with B integral over A. Let $\mathfrak{p} \subseteq B$ be prime and $\mathfrak{q} = \mathfrak{p}^c = \mathfrak{p} \cap A$ be prime in A. Then $\mathfrak{p}$ is maximal if and only if $\mathfrak{q}$ is also maximal.

Proof. $B/\mathfrak{q}$ is integral over $A/\mathfrak{a}$ and both are integral domains since $\mathfrak{p}, \mathfrak{q}$ are prime. By the lemma, if one is a field (so the ideal is maximal) then the other must also be a field. □

It is important that B be integral over A, without this condition the result may not always be true.

Example. As a simple example take $\mathbb{Z} \subseteq \mathbb{Q}$ and note that $\mathbb{Q}$ is not integral over $\mathbb{Z}$ ($\mathbb{Z}$ is integrally closed). The preimage of $(0) \subseteq \mathbb{Q}$, which is maximal, is $(0) \subseteq \mathbb{Z}$, which is not.

Proposition 7.25. Let $A \subseteq B$ be rings with B integral over A. Let $\mathfrak{q} \subseteq \mathfrak{q}'$ be prime ideals in B such that $\mathfrak{q} \cap A = \mathfrak{q}' \cap A$, then $\mathfrak{q} = \mathfrak{q}'$.

Proof. Consider the multiplicative set $S = A \setminus \mathfrak{p}$, then $S^{-1}B$ is integral over $S^{-1}A$. Since these are localizations, let $\mathfrak{m}$ be the unique maximal ideal of $S^{-1}A$. Now since localization is exact

$$\mathfrak{n} = S^{-1}\mathfrak{q} \subseteq S^{-1}\mathfrak{q}' = \mathfrak{n}'$$

But the contractions of $\mathfrak{n}, \mathfrak{n}'$ are both $\mathfrak{m}$, which means they are both maximal and thus equal, therefore $\mathfrak{q} = \mathfrak{q}'$. □

This proposition essentially states that if two ideals contract to the same ideal then they must also be the same. This feature along with previous properties we proved are very desirable to have, hence the idea of integral dependence making "good" rings.

Theorem 7.26. Let $A \subseteq B$ be rings with B integral over A. If $\mathfrak{p} \subseteq A$ is prime, then there exists a prime ideal $\mathfrak{q} \subseteq B$ such that $\mathfrak{q} \cap A = \mathfrak{p}$.

Proof. Since $B_\mathfrak{p}$ is integral over $A_\mathfrak{p}$, we can form the commutative diagram

$$\begin{array}{ccc} A & \hookrightarrow & B \\ {\scriptstyle\alpha}\downarrow & & \downarrow{\scriptstyle\beta} \\ A_\mathfrak{p} & \hookrightarrow & B_\mathfrak{p} \end{array}$$

Since $B_\mathfrak{p}$ may not be local, let $\mathfrak{n}$ be any maximal ideal, then $\mathfrak{m} = \mathfrak{n} \cap A_\mathfrak{p}$ is maximal. Furthermore it is the unique maximal ideal of $A_\mathfrak{p}$ since it is a local ring. If we take $\mathfrak{q} = \beta^{-1}(\mathfrak{n})$, then we get a prime ideal such that $\mathfrak{q} \cap A = \alpha^{-1}(\mathfrak{m}) = \mathfrak{p}$. □

Theorem 7.27 (Going Up)**.** Let $A \subseteq B$ be rings with B integral over A. Suppose we have two chains of prime ideals

$$\mathfrak{p}_1 \subseteq \mathfrak{p}_2 \subseteq \cdots \subseteq \mathfrak{p}_n \subseteq A$$
$$\mathfrak{q}_1 \subseteq \mathfrak{q}_2 \subseteq \cdots \subseteq \mathfrak{q}_m \subseteq B$$

with $m < n$ and where $\mathfrak{q}_i \cap A = \mathfrak{p}_i$ for $1 \leq i \leq m$. Then this chain can be extended to $\mathfrak{q}_1 \subseteq \cdots \subseteq \mathfrak{q}_n$ such that the condition holds for all $1 \leq i \leq n$.

Proof. By induction this can be reduced to the case of $m = 1, n = 2$. Let $\overline{A} = A/\mathfrak{p}_1$ and $\overline{B} = B/\mathfrak{q}_1$, then $\overline{A} \subseteq \overline{B}$ with $\overline{B}$ integral over $\overline{A}$. Thus there exists a prime ideal $\mathfrak{q}_2 \subseteq \overline{B}$ such that $\overline{\mathfrak{q}_2} \cap \overline{A} = \overline{\mathfrak{p}_2}$ (the image of $\mathfrak{p}_2$ in $\overline{A}$). Lifting this back to B gives us a prime ideal $\mathfrak{q}_2$ which satisfies our properties. □

For completeness we will also list (but not prove) the going down theorem. The names going up and going down refer to the direction in which we extend inclusions of ideals.

Theorem 7.28 (Going Down)**.** Let $A \subseteq B$ be rings with B integral over A. Suppose we have two chains of prime ideals

$$A \supseteq \mathfrak{p}_1 \supseteq \mathfrak{p}_2 \supseteq \cdots \supseteq \mathfrak{p}_n$$
$$B \supseteq \mathfrak{q}_1 \supseteq \mathfrak{q}_2 \supseteq \cdots \supseteq \mathfrak{q}_m$$

with $m < n$ and where $\mathfrak{q}_i \cap A = \mathfrak{p}_i$ for $1 \leq i \leq m$. Then this chain can be extended to $\mathfrak{q}_1 \supseteq \cdots \supseteq \mathfrak{q}_n$ such that the condition holds for all $1 \leq i \leq n$.

7.5 Valuation Rings

Valuation rings, as the name suggests, are rings with a valuation: an associated function which assigns a "size" to elements of the ring. Here we will define valuation rings in two ways: the obvious way using a valuation and the traditional way.

Definition. *A discrete valuation on a field k is a function*

$$v : k \to \mathbb{Z} \cup \{\infty\}$$

such that

1. $v(0) = \infty$
2. $v(xy) = v(x) + v(y)$
3. $v(x + y) \geq min\{v(x), v(y)\}$

Here are some examples of valuations on familiar fields.

Example. On $\mathbb{R}$ we can define $v(x) = -\log|x|$, this is the Archimedian valuation, note however that it is not discrete since it takes real values.

Example. On $\mathbb{Q}$ for some prime p we can define

$$v(x) = -(\text{large power of } p \text{ which divides } x)$$

if $x = m/n$, then define $v(x) = v(m) - v(n)$. This is the p-adic valuation.

Example. On $\mathbb{C}[x]$ for some $a \in \mathbb{C}$ we can define $v(f/g) = v(f) - v(g)$. There are two valuations

$$v_a(f) = \text{largest power of } x - a \text{ which divides } f$$

this is the order of vanishing at a. The second valuation is the order of vanishing at infinity

$$v_\infty(f) = \deg f$$

Definition. *If v is a (discrete) valuation, then the valuation ring of v is*

$$A = \{x \in k \mid v(x) \geq 0\}$$

which has a unique maximal ideal

$$\mathfrak{m} = \{x \in k \mid v(x) > 0\}$$

Let's use one of the valuations we created earlier and find its valuation ring.

Example. Take $a = 0$ and consider the ring $k[x]$ with valuation v_0

$$\begin{aligned} A &= \left\{ \frac{f}{g} \middle| v_0\left(\frac{f}{g}\right) \geq 0 \right\} \\ &= \left\{ \frac{f}{g} \middle| x \nmid g \right\} \\ &= k[x]_{(x)} \end{aligned}$$

Observe that for any $x \in k$ there are four possibilities for its value

$$v(x) \begin{cases} < 0 & \text{if } x \notin A \text{ and } x^{-1} \in A \\ = 0 & \text{if } x \in A \text{ is a unit} \\ > 0 & \text{if } x \in A \text{ is not a unit} \\ = \infty & \text{if } x = 0 \end{cases}$$

which we can see from some quick calculations

$$v(1 \cdot 1) = v(1) + v(1) = v(1) \to v(1) = 0$$
$$\therefore v(x^{-1}) = -v(x)$$

Using these facts, we can define valuations in another way: the traditional way. [7]

Definition. *Let k be a field, a subring $B \subseteq k$ is a valuation ring of k if for all nonzero $x \in k$ either $x \in B$ or $x^{-1} \in B$.*

Proposition 7.29. Note the following properties of valuation rings:

1. B is local
2. If $B \subseteq B'$, then B' is also a valuation ring
3. B is integrally closed

Proof. 1) Let $\mathfrak{m}$ be the set of non-units of B so that

$$x \in \mathfrak{m} \iff x = 0 \text{ or } x^{-1} \notin B$$

If $a \in B$ and $x \in \mathfrak{m}$, then ax is not a unit if $x \neq 0$ since otherwise $(ax)^{-1} \in B$ and so $x^{-1} = a \cdot (ax)^{-1} \in B$. If $x, y \in \mathfrak{m}$ are nonzero then either xy^{-1} or $x^{-1}y$ are in B. Suppose without loss of generality that $xy^{-1} \in B$, then

$$x + y = (1 + xy^{-1})y \in B\mathfrak{m} \subseteq \mathfrak{m}$$

Thus $\mathfrak{m}$ is an ideal and B is local.

2) Obvious from definitions

3) Let $x \in k$ be integral over B and suppose $x^{-1} \in B$ (if $x \in B$, then we are already done). From the integral condition we have

$$x^n + b_1 x^{n-1} + \cdots + b_n = 0$$

multiplying by $x^{-(n-1)} \in B$, we get

$$x + b_1 + \cdots + b_n x^{-(n-1)} = 0$$
$$\therefore x = -(b_1 + b_2 x^{-1} + \cdots + b_n x^{-(n-1)}$$
$$\therefore x \in B$$

Thus B is integrally closed. □

[7] By traditional, I mean it was used in the book

Let k be a field and Σ be the set of pairs (A, f) where for some algebraically closed field Ω, $A \subseteq k$ is a subring and $f : A \to \Omega$ is a homomorphism. Define the partial order

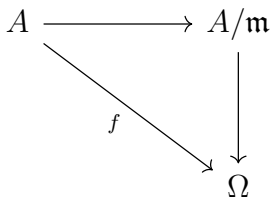
$$(A, f) \leq (A', f') \iff A \subseteq A',\ f'|_A = f$$

which gives us a maximal element of Σ by Zorn's lemma. Let (B, g) be this maximal element, we will show that B is a valuation ring of k.

There is still the question of how we choose Ω. Typically we start with a ring $A \subseteq k$ and let $\mathfrak{m}$ be a maximal ideal of A. The quotient will be a field so we choose

$$\Omega = \overline{A/\mathfrak{m}}$$

This gives us a pair $(A, f) \in \Sigma$ where



To prove this we must first establish some lemmas.

Lemma 7.30. B is local with unique maximal ideal $\mathfrak{m} = \ker g$.

Proof. Since we have a map $g : B \to \Omega$, $\ker g = g^{-1}((0))$ is a prime ideal, let's call it $\mathfrak{p}$. Consider the ring

$$B_{\mathfrak{p}} = \left\{ \frac{x}{y} \middle| x \in B, y \in B \setminus \mathfrak{p} \right\}$$

We can extend the map g to get

$$\overline{g} : B_{\mathfrak{p}} \to \Omega \qquad \frac{x}{y} \mapsto \frac{g(x)}{g(y)}$$

Note that this works since Ω is a field, this gives

$$(B_{\mathfrak{p}}, \overline{g}) \geq (B, g)$$

which is a contradiction because (B, g) is already the maximal element. Thus it must be that $B = B_{\mathfrak{p}}$ is local. □

Lemma 7.31. Let $x \in k$ be nonzero, then

$$\mathfrak{m}[x] \subset B[x] \quad \text{or} \quad \mathfrak{m}[x^{-1}] \subset B[x^{-1}]$$

Proof. Suppose that we have equality in both cases. Since $1 \in B[x], B[x^{-1}]$, we can write

$$u_0 + u_1 x + \cdots + u_m x^m = 1 \quad u_i \in \mathfrak{m}$$
$$v_0 + v_1 x^{-1} + \cdots + v_n x^{-n} = 1 \quad v_j \in \mathfrak{m}$$

in which we assume the degrees m, n are as small as possible. Furthermore suppose $m \geq n$ and multiply through by x^n to get

$$(1 - v_0)x^n = v_1 x^{n-1} + \cdots + v_n$$

Since $v_0 \in \mathfrak{m}$, this implies that $1 - v_0$ is a unit and so we can replace $x^n, x^{n+1}, \ldots, x^m$ with expressions in $x^{n-1}, \ldots, 1$. But this implies our first equation wasn't minimal, a contradiction. Thus at least one of the inclusions must be strict. □

Now we are ready to prove the theorem.

Theorem 7.32. If $(B, g) \in \Sigma$ is maximal, then B is a valuation ring.

Proof. Let $x \in k$ be nonzero, we must show that either $x \in B$ or $x^{-1} \in B$. Consider $B[x]$ and $B[x^{-1}]$ and suppose without loss of generality that (from the previous lemma)

$$\mathfrak{m}[x] \neq B[x]$$

Thus $\mathfrak{m}[x]$ is contained within some maximal ideal $\mathfrak{m}'$ of $B[x]$. Note that

1. $\mathfrak{m}' \cap B$ is an ideal of B
2. $1 \notin \mathfrak{m}'$ (since $1 \notin \mathfrak{m}$, so its contraction is a proper ideal
3. $\mathfrak{m}' \cap B \subseteq \mathfrak{m}$
4. $\mathfrak{m}$ is maximal
5. Therefore $\mathfrak{m}' \cap B = \mathfrak{m}$

Using this we have the inclusions [8]

$$\begin{array}{ccc} B & \hookrightarrow & B' = B[x] \\ \uparrow & & \uparrow \\ \mathfrak{m}' \cap B = \mathfrak{m} & \hookrightarrow & \mathfrak{m}' \end{array}$$

which induces a map

$$k = B/\mathfrak{m} \hookrightarrow B'/\mathfrak{m}' = k'$$

This is an embedding of fields and we claim k' is an algebraic extension of k. If we denote the image of x in k' with $\overline{x}$, then

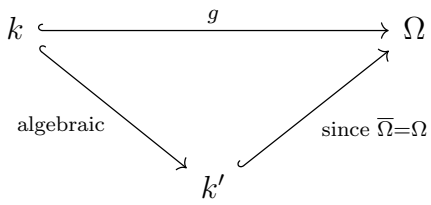
$$k' = k[\overline{x}]$$

which means $\overline{x}$ is algebraic and thus so is k'. Recalll that g also induces the embedding

$$\overline{g} : k = B/\mathfrak{m} \hookrightarrow \Omega$$

[8] Denoted by the curved arrows, these arrows typically denote injective maps but we can think of inclusion as a trivial map so this makes sense

This can be extended with



Thus we have another pair (B', g') where

$$g' : B' \longrightarrow k' = B'/\mathfrak{m}' \hookrightarrow \Omega$$

But (B, g) is already maximal, so we must have $B' = B[x] = B$ which implies $x \in B$. □

Corollary 7.33. Let $A \subseteq k$ be a subring, then the integral closure $\overline{A}$ is the intersection of all valuation rings containing k.

Proof. Let B be a valuation ring with $A \subseteq B$, since B is integrally closed we have $\overline{A} \subseteq B$.

For the other inclusion, suppose $x \notin \overline{A} \to x \notin A' = A[x^{-1}]$ so x^{-1} is a non-unit in A and is included in some maximal ideal $\mathfrak{m}'$. Let Ω be the algebraic closure of the field $k' = A'/\mathfrak{m}'$. The restriction of the natural map

$$\pi : A' \to k'$$

to A gives a homomorphism

$$\pi|_A : A \to \Omega$$

This can be extended to some valuation ring $B \supseteq A$. Since x^{-1} maps to 0, we get that $x \notin B$ and so $B \subseteq \overline{A}$, which establishes the equality. □

Valuation rings can also be used to prove one form of Hilbert's Nullstellensatz ("theorem of zeros") which I will list but not prove.

Theorem 7.34 (Nullstellensatz)**.** Let k be a field and B a finitely generated k-algebra. If B is a field then it is a finite algebraic extension of k.

7.6 Chain Conditions

We've discussed chain conditions previously with the introduction of Noetherian and Artinian modules. Here we will reintroduce them and discuss their properties a bit more. Recall that the reason we care about these modules is that in general finitely generated modules are much nicer to work with than arbitrary modules (think of them as similar to finite dimensional vector spaces). The issue is that submodules of finitely generated modules are generated not finitely generated.

Example. Consider $A = k[x_1, x_2, \ldots]$, take $M = A$ and $N = (x_1, x_2, \ldots)$. As A-modules, M is finitely generated ($M = (1)$), but $N \subseteq M$ is not.

Essentially we want a notion of "nice" rings A where this doesn't occur for A-modules.

Definition. *A partially ordered set $(\Sigma, \leq)$ is said to satisfy the ascending chain condition if for any chain*

$$A_1 \leq A_2 \leq \cdots \leq A_n \leq \cdots$$

there exists some $k \geq 0$ such that

$$A_k = A_{k+1} = \cdots$$

In other words, the chain cannot be infinite and must terminate at some point.

The descending chain condition is described the same way, with reversed inclusions.

Definition. *Let A be a ring and M an A-module. M is Noetherian if it satisfies the ascending chain condition on submodules and Artinian if it satisfies the descending chain condition.*

Example. $\mathbb{Z}$ is Noetherian but not Artinian. If we have the chain

$$(m_1) \subseteq (m_2) \subseteq (m_3) \subseteq \cdots$$

then we also have

$$m_1 \mid m_2 \mid m_3 \mid \cdots$$

which must stabilize. However in the other direction we have

$$(1) \supseteq (2) \supseteq (4) \supseteq (8) \supseteq \cdots$$

Example. A field is both Noetherian and Artinian

We say a ring is Noetherian/Artinian if it is Noetherian/Artinian as a module over itself. In other words if they satisfy the ascending/descending chain condition on their ideals. Now we prove the main reason why we care about Noetherian modules.

Theorem 7.35. A module M is Noetherian if and only if every submodule $N \subseteq M$ if finitely generated

Proof. Given earlier: see proposition 6.7 □

Now we will look at Noetherian/Artinian modules under certain operations.

Proposition 7.36. Suppose we have a short exact sequence

$$0 \longrightarrow M' \longrightarrow M \longrightarrow M'' \longrightarrow 0$$

then M is Noetherian/Artinian if and only if both M', M'' are Noetherian/Artinian

Proof. We will prove this for Noetherian modules, the Artinian case is similar. Submodules of M', M'' can be identifies with submodules of M, preserving order. Thus if M is Noetherian then so must be M', M''.

Conversely suppose M', M' are Noetherian. Let $N \subseteq M$ be a submodule and consider its image $\overline{N} \subseteq M''$, which is finitely generated by assumption. Note

$$\overline{N} = N/N \cap M' \subseteq M''$$

Let $\overline{n}_1, \ldots, \overline{n}_k$ generate $\overline{N}$, we can lift these to elements $n_1, \ldots, n_k \in N$. The submodule $N' = N \cap M' \subseteq M'$ is also finitely generated, let $p_1, \ldots, p_r$ be its generators. We claim that $\{n_1, \ldots, n_k, p_1, \ldots, p_r\}$ generate N. Consider some $x \in N$, then

$$\begin{aligned} \overline{x} &= a_1\overline{n}_1 + \cdots a_k\overline{n}_k \\ \therefore \overline{x - a_1n_1 - \cdots - a_kn_k} &= 0 \\ \therefore x - a_1n_1 - \cdots - a_kn_k &\in N' \\ \therefore x - a_1n_1 - \cdots - a_kn_k &= b_1p_1 + \cdots + b_rp_r \\ \therefore x &= a_1n_1 + \cdots + a_kn_k + b_1p_1 + \cdots + b_rp_r \end{aligned}$$

and thus x is finitely generated by our set. □

Corollary 7.37. The direct product of Noetherian/Artinian modules is stays Noetherian/Artinian.

Proof. Use induction and the proposition on the exact sequence

$$0 \longrightarrow M_n \longrightarrow \bigoplus_{i=1}^{n} M_i \longrightarrow \bigoplus_{i=1}^{n-1} M_i \longrightarrow 0$$

□

Corollary 7.38. Let A be Noetherian/Artinian and $\mathfrak{a} \subseteq A$ an ideal. Then $A/\mathfrak{a}$ is a Noetherian/Artinian ring.

Proof. Apply the proposition to the exact sequence

$$0 \longrightarrow \mathfrak{a} \longrightarrow A \longrightarrow A/\mathfrak{a} \longrightarrow 0$$

□

Corollary 7.39. If A is Noetherian/Artinian, then any finitely generated A-module is also Noetherian/Artinian.

Proof. A finitely generated module is a quotient of A^n for some n and thus is Noetherian/Artinian by previous results. □

This last corollary reduced the problem of finding Noetherian modules to finding Noetherian rings. The next theorem is very useful in helping us generate more Noetherian rings.

Theorem 7.40 (Hilbert Basis Theorem)**.** If A is a Noetherian ring then $A[x]$ is also Noetherian.

Proof. Let A be Noetherian, we need to show that any $A[x]$-submodule of $A[x]$ is finitely generated. In other words we must show that any ideal of $A[x]$ is finitely generated. Let $\mathfrak{a} \subseteq A[x]$ be any ideal and define

$$\mathfrak{a}' = \{l.c.(f) \mid f \in \mathfrak{a}\} \subseteq A$$

where $l.c.$ denotes the leading coefficient, e.g.

$$l.c.(a_0x^n + a_1x^{n-1} + \cdots + a_n) = a_0$$

It's clear that $\mathfrak{a}'$ is an ideal of A, since A is Noetherian suppose $\mathfrak{a}'$ is generated by $t_1, \ldots, t_n$. Pick some polynomials $f_1, \ldots, f_n$ such that $l.c.(f_i) = t_i$ and define

$$k = max\,\{\deg\, f_i\} \in \mathbb{Z}$$

We claim that every $f \in \mathfrak{a}$ can be written as

$$f = g_1f_1 + g_2f_-2 + \cdots + g_nf_n + h$$

where $g_i \in A[x]$ and $\deg\, h < k$. To see this we can look at the leading coefficients of f (note $\deg\, f \geq k$)

$$l.c.(f) \in \mathfrak{a}' \quad \rightarrow \quad l.c.(f) = b_1t_1 + \cdots + b_nt_n$$

this allows us to take

$$g_1 = b_1x^{\deg\, f - \deg\, f_1}$$

$$\vdots$$

$$g_n = b_nx^{\deg\, f - \deg\, f_n}$$

With this choice of g_i, note

$$l.c.(g_1f_1 + \cdots g_nf_n) = l.c.(f)$$

with the same degrees on leading coefficients, thus

$$\deg\,(f - g_1f_1 - \cdots - g_nf_n) < \deg\, f$$

We can reduce this until the we get a degree less than k. Let T be the A-submodule of $A[x]$ generated by $\left\{1, x, x^2, \ldots, x^{k-1}\right\}$. We've just proved

$$\mathfrak{a} = (f_1, \ldots, f_n) + \mathfrak{a} \cap T$$

But T is finitely generated and A is Noetherian so $\mathfrak{a} \cap T$ is also finitely generated. Thus $\mathfrak{a}$ is finitely generated as an $A[x]$-module. □

Corollary 7.41. If A is Noetherian, then so is $A[x_1, \ldots, x_n]$

Proof. Obvious from induction □

Corollary 7.42. If A is Noetherian and B a finitely generated A-algebra, then B is Noetherian.

Proof. If B is finitely generated then $B = A[x_1, \ldots, x_n]/I$ for some ideal I. This gives us a short exact sequence

$$0 \longrightarrow I \longrightarrow A[x_1, \ldots, x_n] \longrightarrow B \longrightarrow 0$$

B is Noetherian because $A[x_1, \ldots, x_n]$ is Noetherian. □

This last corollary implies, in particular, that every finitely generated ring and every finitely generated algebra over a field is Noetherian.

Part IV
Theory of Field Extensions

8 Field Theory

8.1 Introduction to Field Extensions

Recall that a field is a commutative ring with identity such that every nonzero element in invertible.

Definition. *The characteristic of a field F, denoted $ch(F)$ is the smallest positive integer p such that $p \cdot 1 = 0$. If no such p exists then $ch(F) = 0$.*

The characteristic is necessarily prime, if $p = mn$ is the characteristic then

$$mn \cdot 1 = (m \cdot 1)(n \cdot 1) = 0$$

since zero divisors don't exist in a field $ch(F) = min(m, n)$. This also implies that if $n \cdot 1 = 0$ then $p \mid n$.

Proposition 8.1. Suppose $ch(F) = p$, then $p \cdot a = 0$ for any $a \in F$

Proof. $p \cdot a = p \cdot 1a = (p \cdot 1)a = 0$ □

So the characteristic can be thought of as the number of times we must multiply an element in order to reach 0.

Example. $ch(\mathbb{Q}) = ch(\mathbb{R}) = 0$. For a finite field $ch(F_p) = p$.

Suppose we define $(-n) \cdot 1 = -(n \cdot 1)$ and $0 \cdot 1 = 0$, then we can define a map

$$\phi : \mathbb{Z} \to F \qquad n \mapsto n \cdot 1$$

Then $\ker \phi = ch(F)\mathbb{Z} = p\mathbb{Z}$ which gives an injection

$$\mathbb{Z}/p\mathbb{Z} \to F$$

If $p = 0$ then there is a subfield of F isomorphic to $\mathbb{Q}$ otherwise the subfield is isomorphic to $\mathbb{Z}/p\mathbb{Z}$.

Definition. *Let F be a field, the subfield generated by $1 \in F$ is called the prime subfield of F and is isomorphic to either $\mathbb{Q}$ if $ch(F) = 0$ or $\mathbb{F}_p$ if $ch(F) = p$.*

Example. It's easy to see then that the prime subfields of $\mathbb{Q}$ and $\mathbb{R}$ is $\mathbb{Q}$

Note that for a field of characteristic p, the "freshman's dream" is true

$$(a + b)^p = a^p + b^p$$

As a consequence we can define the Frobenius endomorphism

$$F : k \to k \qquad x \mapsto x^p$$

which is a valid endomorphism only in a field of characteristic p.

Definition. *If K is a field containing the subfield F then K is an extension field of F, denoted K/F (read "K over F") or by the diagram*

$$\begin{array}{c} K \\ | \\ F \end{array}$$

The field F is called the base field.

If K/F is any extension of fields, then multiplication in K makes it a vector space over F. Since any field is an extension of its prime subfield they can be considered as a vector space over its prime field. Using this train of thought, the dimension of a field considered as a vector space has some significance.

Definition. *The degree (or relative degree or index) of a field extension K/F, denoted $[K : F]$ is the dimension of K as a vector space over F, that is*

$$[K : F] = \dim_F K$$

An extension is finite if it has finite degree and is infinite otherwise.

A fundamental question we can ask arises when trying to do basic algebra in certain fields. For instance in the field of real numbers $\mathbb{R}$, the polynomial $x^2+1 = 0$ has no solutions. Thus we can ask if we can construct a field extension in which this is possible and indeed we can, we just need to append i to get $\mathbb{C}$. For a general polynomial $p(x) \in F[x]$ which is irreducible (i.e. has no roots in F), we would like to know if we can construct a field extension K/F which contains a root of $p(x)$. The answer is yes and our work on this problem will lead into Galois theory.

Lemma 8.2. Let $\phi : F \to F'$ be a homomorphism of fields, then ϕ is either identically zero or injective.

Proof. $\ker \phi$ is an ideal of F, but the only ideals of a field are 0 and F. Thus either ϕ is zero ($\ker \phi = F$) or injective ($\ker \phi = 0$). □

Theorem 8.3. Let F be a field and $p(x) \in F[x]$ be an irreducible polynomial, then there exists a field extension K/F in which $p(x)$ is reducible.

Proof. Consider the quotient

$$K = F[x]/(p(x))$$

Since $p(x)$ is irreducible, the ideal $(p(x))$ is maximal which makes K a field. We have a canonical map

$$\pi : F[x] \to F[x]/(p(x)) \qquad f \mapsto f + (p(x))$$

If we restrict π to $F \subset F[x]$, then we get a homomorphism

$$\varphi = \pi|_F : F \to K$$

Since $\varphi(1) = 1 + (p(x))$, φ is not identically zero. So by the lemma im $\varphi \cong F$ which is a copy of F in K, implying that F is a subfield of K. If $\pi(x) = \overline{x}$ then

$$p(\overline{x}) = \overline{p(x)} = p(x) \mod (p(x)) = 0$$

and so $\overline{x}$ is a root of $p(x)$ in K. □

In the proof we constructed an extension by first forming the quotient, but what exactly is this quotient. To further understand what K is it is useful to ask for a basis since K is a vector space over F.

Theorem 8.4. Let $p(x) \in F[x]$ be an irreducible polynomial of degree n over a field F. Let $K = F[x]/(p(x))$ and $\theta = x \mod (p(x)) \in K$, then $1, \theta, \theta^2, \ldots, \theta^{n-1}$ for a basis of K which implies $[K : F] = n$. Hence

$$K = \left\{ a_0 + a_1\theta + a_2\theta^2 + \cdots + a_{n-1}\theta^{n-1} \mid a_i \in F \right\}$$

which is the set of polynomials of degree less than n.

Proof. Let $a(x) \in F[x]$ be some polynomial. We can perform the division algorithm to get

$$a(x) = q(x)p(x) + r(x) \qquad q(x), r(x) \in F[x] \quad deg(r) < n$$

It thus follows that $a(x) = r(x) \mod (p(x))$, proving that every residue class of $F[x]/(p(x))$ can be represented by a polynomial of degree less than n. Thus the images

$$1, x, x^2, \ldots, x^{n-1} \to 1, \theta, \theta^2, \ldots, \theta^{n-1}$$

span K as a vector space over F. What remains is to show they are linearly independent. Suppose they are dependent, that is

$$b_0 + b_1\theta + \cdots + b_{n-1}\theta^{n-1} = 0$$

where $b_i \in F$ are not all zero. This is equivalent to

$$b_0 + b_1\theta + \cdots + b_{n-1}\theta^{n-1} \equiv 0 \mod (p(x))$$

which would imply that $p(x)$ divides the left hands side which is impossible because it is of degree less than n. Thus $1, \theta, \ldots, \theta^{n-1}$ form a basis for K over F. □

Now that we have a description of K we want to know how elements interact in the field. Addition is clearly just normal polynomial addition but multiplication is a bit more involved. We want to be able to multiply two polynomials of degree $< n$ and obtain another polynomial of degree $< n$. In other words we want the representative for the coset $f(x)g(x) + (p(x))$. This is done by dividing fg by p and the remainder will be the representative of the coset. To summarize...

Corollary 8.5. Let $a(\theta), b(\theta) \in K$. Addition is standard polynomial addition and multiplication is

$$a(\theta)b(\theta) = r(\theta) = a(\theta)b(\theta) \mod p(x)$$

These results make K into a field, which of course makes sense given that it is a *field* extension.

Example. Consider the polynomial $x^2+1 \in \mathbb{R}[x]$. The extension if thus

$$K = \mathbb{R}[x]/(x^2+1) \qquad [k : \mathbb{R}[x]] = deg(x^2+1) = 2$$

If we let $\theta = x \mod (x^2+1)$, then

$$K = \{a + b\theta \mid a, b, \in \mathbb{R}\}$$

Note that this also implies $x^2 \mod (x^2+1) = -1 \to \theta^2 = -1$ and that θ satisfies

$$(a+b\theta) + (c+d\theta) = (a+c) + (b+d)\theta$$
$$(a+b\theta)(c+d\theta) = ac + (bc+ad)\theta + bc\theta^2 = (ac-bd) + (bd+ac)\theta$$

This should be familiar because we can identify $\theta = i$ and thus we've construct $K = \mathbb{C}$. This gives us an isomorphism

$$\mathbb{C} \cong \mathbb{R}[x]/(x^2+1)$$

but this isomorphism is not canonical.

A useful way to determine if polynomials are irreducible is Eisenstein's criterion.

Proposition 8.6 (Eisenstein)**.** Let $f = a_n x^n + \cdots + a_0 \in \mathbb{Z}[x]$, f is irreducible over $\mathbb{Q}$ (and $\mathbb{Z}$) if there is exists a prime p such that

1. $p \mid a_i$ for all $i < n$
2. $p \nmid a_n$
3. $p^2 \nmid a_0$

A more general form exists for all integral domains R, but for our purposes this will suffice.

Example. Let $k = \mathbb{Q}$, $f(x) = x^3 - -3x + 1$ which is irreducible over k, and take the extension $L = k/(f)$. The extension has as a basis $\{1, x, x^2\}$. When doing computations we must reduce $\mod f$ which is the same thing as saying $x^3 - 3x + 1 = 0 \to x^3 = 3x - 1$. For example take

$$a = x^2 + 2x = (0, 2, 1) \in L$$
$$b = x^2 + 1 = (1, 0, 1) \in L$$

If we multiply the two we get

$$\begin{aligned} ab &= (x^2+2x)(x^2+1) \\ &= x^4 + 2x^3 + x^2 + 2x \\ &= x(3x-1) + 2(3x-1) + x^2 + 2x \\ &= 3x^2 - x + 6x - 2 + x^2 + 2x \\ &= 4x^2 + 7x - 2 \\ \therefore (0,2,1) \cdot (1,0,1) &= (-2, 7, 4) \in L \end{aligned}$$

Now let's do a more adventurous example where we work in a less familiar field.

Example. Take $k = \mathbb{Z}/2\mathbb{Z}$ and $f(x) = x^2 + 2x + 1$. We will show that the extension is

$$L = k[x]/(f) \cong \mathbb{F}_4$$

Note that the mod relation here is $x^2 = -x - 1 = 1 + x$ since we are in $\mathbb{Z}/2\mathbb{Z}$, this implies that our basis is $\{0, 1, x, 1 + x\}$. The full multiplication table is.

	0	1	x	1+x
0	0	0	0	0
1	0	1	x	1+x
x	0	x	1+x	1
1+x	0	1+x	1	x

This happens to be how we construct a field with four elements.

When we discuss extensions in which a root for some polynomial exists, generally we would like to find the smallest such field.

Definition. *Let K/F be an extension and $\alpha, \beta, \ldots \in K$. Then the smallest subfield of K containing α, β, ldots is called the field generated by $\alpha, \beta, \ldots$ over F and is denoted $F(\alpha, \beta, \ldots)$*

Definition. *If a field K is generated by a single element α over F, that is $K = F(\alpha)$, then it is called a simple extension of F and α is called a primitive element for the extension.*

When α is a root of some irreducible polynomial, then there is a connection between $F(\alpha)$ and the field extension constructed in theorem 8.3.

Theorem 8.7. Let F be a field and $p(x) \in F[x]$ be an irreducible polynomial. Suppose K/F is an extension containing a root α of $p(x)$ and $F(\alpha)$ be the subfield of K generated by α over F, then

$$F(\alpha) \cong F[x]/(p(x))$$

Proof. There is a natural homomorphism

$$\phi : F[x] \to F(a) \subseteq K \qquad a(x) \mapsto a(\alpha)$$

Since $p(\alpha) = 0$ by assumption, $p(x) \in \ker \phi$ which induces a homomorphism

$$\phi : F[x]/(p(x)) \to F(\alpha)$$

Since $p(x)$ is irreducible, this is a homomorphism of fields and since ϕ is definitely not zero (it's the identity on F), this is an isomorphism of fields with the image im ϕ. Since the image is a subfield of $F(\alpha)$ containing α by definition it must be $F(\alpha)$ itself so the map is surjective and thus an isomorphism. □

The difference between this theorem and theorem 8.3 is that this theorem assumes the existence of a root α in a field K whereas we prove that K exists in 8.3. Instead we prove that the extension constructed is isomorphic to the subfield generated by α $F(\alpha)$.

Corollary 8.8. If $p(x)$ is degree n, then it follows that the generated extension is

$$F(\alpha) = \left\{a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} \mid a_i \in F\right\} \subseteq K$$

Example. $\mathbb{Q}(\sqrt[3]{2}) = \left\{a + b\sqrt[3]{2} + c\sqrt[3]{4} \mid a, b, c \in \mathbb{Q}\right\}$

The theorem implies that all fields obtaining by adjoining a root of a polynomial are isomorphic which means that the roots of a polynomial are algebraically indistinguishable. The property that different roots of the same irreducible polynomial have the same algebraic properties can be extended.

Theorem 8.9. Let $\varphi : F \to F'$ be an isomorphism of fields and $p(x) \in F[x]$, $p'(x) \in F'[x]$ be irreducible polynomials where $p'(x)$ is obtained by applying φ to $p(x)$. Let α be a root of $p(x)$ and β a root of $p'(x)$, then there is an isomorphism

$$\sigma : F(a) \to F'(\beta) \qquad \alpha \mapsto \beta$$

which extends φ (i.e. σ restricted to F is φ).

Proof. φ induced a natural isomorphism between $F[x]$ and $F'[x]$ which maps the maximal ideal $(p(x))$ to $(p'(x))$. We can quotient using these ideals to get another isomorphism

$$F(\alpha) \cong F[x]/(p(x)) \to F'[x]/(p'(x)) \cong F'(\beta)$$

Composing these isomorphisms gives us σ and it's clear that restricting σ to F gives the original isomorphism φ. □

This theorem can also be stated pictorially.

$$\begin{array}{ccc} F(\alpha) & \xrightarrow{\quad\sigma\quad} & F'(\beta) \\ \Big| & & \Big| \\ F & \xrightarrow{\quad\varphi\quad} & F' \end{array}$$

where σ, φ are field isomorphisms.

8.2 Algebraic Extensions

Definition. *Let K be a field extension of a field F. An element $\alpha \in K$ is algebraic over F if it is the root of some polynomial $p(x) \in F[x]$, otherwise it is transcendental over F. The extension K is algebraic if every element $\alpha \in K$ is algebraic over F.*

If some element α is algebraic over F then it will be algebraic over any extension field because any $f(x)$ with α as a root will necessarily have coefficients in L.

Proposition 8.10. Let α be algebraic over F, then there is a unique monic irreducible polynomial $m_{\alpha,F}(x) \in F[x]$ which has α as a root. A polynomial $f(x)$ has α as a root if and only if $m\alpha, F \mid f(x)$

Proof. Let $g(x) \in F[x]$ be a polynomial of minimal degree which has α as a root. We can multiply g by a constant to make it monic. If g is reducible, that is $g(x) = a(x)b(x)$, then

$$g(\alpha) = a(\alpha)b(\alpha) = 0 \to a(\alpha) = 0 \text{ or } b(\alpha) = 0$$

But this is a contradiction so g is irreducible. Suppose $f(x) \in F[x]$ also has α as a root, then

$$f(x) = q(x)g(x) + r(x) \qquad deg(r(x)) < deg(g(x))$$

This implies that

$$f(\alpha) = q(\alpha)g(\alpha) + r(\alpha) = r(\alpha) = 0$$

which is a contradiction because g is of minimal degree, so $r(x) = 0$. Thus $g \mid f$, in particular it would divide any other irreducible polynomial with α as a root if they exist, therefore g is also unique. □

Corollary 8.11. If L/F is a field extension and α is algebraic over both F and L, then $m_{\alpha,L}$ divides $m_{\alpha,F}$

Proof. This follows from the proposition because $m_{\alpha,F}$ is a polynomial in L which has α as a root. □

Definition. *This polynomial $m_{\alpha,F}(x)$ is called the minimal polynomial of α. The degree of $m_{\alpha,F}$ is called the degree of α.*

The proposition implies that a monic polynomial with root α is its minimal polynomial if and only if it is irreducible over F. The theorem also has the following consequence

Proposition 8.12. Suppose α is algebraic over F, then

$$F(\alpha) \cong F[x]/(m_\alpha(x)) \qquad [F(\alpha) : F] = deg(m_\alpha) = deg(\alpha)$$

In particular the degree of an algebraic element is the degree of the extension it generates.

Example. For $n > 1$, then minimal polynomial of $\sqrt[n]{2}$ is $x^n - 2$ which is irreducible by Eisenstein. This implies $[\mathbb{Q}(\sqrt[n]{2}) : \mathbb{Q}] = n$.

We can determine if an element is algebraic by investigating the simple extension it generates.

Proposition 8.13. α is algebraic over F if and only if the extension $F(\alpha)/F$ is finite. In other words, if α satisfies a polynomial of degree at most n over F then the degree of $F(\alpha)$ over F is n and vice versa.

Proof. If α is algebraic then

$$[F(\alpha) : F] = deg(m_\alpha) < \infty$$

Thus if α satisfies a polynomial of degree n, then the extension is finite of degree at most n.

Conversely, suppose α is an element of an extension of degree n over F. Then the $n+1$ elements

$$1, \alpha, \alpha^2, \ldots, \alpha^n \in F(\alpha)$$

must be linearly dependent, that is there exists an at most n degree polynomial f such that $f(\alpha) = 0$ and thus α is algebraic. □

This has a powerful corollary.

Corollary 8.14. A finite extension is algebraic.

Proof. Consider some element $\alpha \in K$, the subfield $F(\alpha)$ is a subspace of K over F. Thus

$$[F(\alpha) : F] \leq [K : F] < \infty$$

which means $F(\alpha)/F$ is finite and thus α is algebraic by the proposition. Since this applies for every element, it follows that the extension is algebraic as well. □

The converse is not always true because infinite algebraic extensions do exist, but there is a partial converse which we will see later. For the case of nested subfields such as when L is an extension of K which is an extension of F, their degrees have a convenient relationship.

Theorem 8.15. Let $F \subseteq K \subseteq L$ be fields, then

$$[L : F] = [L : K][K : F]$$

pictorially this is

$$\overbrace{\underbrace{F \quad \subseteq}_{[K:F]} \; K \; \underbrace{\subseteq \quad L}_{[L:K]}}^{[L:F]}$$

If any side of this equation is infinite then so is the other.

Proof. Suppose $[L : K] = m$ and $[K : F] = n$ are finite. Let $\alpha_1, \ldots, \alpha_m$ be a for L over K and $\beta_1, \ldots, \beta_n$ be a basis for K over F. We can write every element in L as

$$a_1\alpha_1 + \cdots + a_m\alpha_m \qquad a_i \in K$$

We can also write every coefficient a_i as

$$a_i = b_{i1}\beta_1 + \cdots + b_{in}\beta_n \qquad b_{ij} \in F$$

In other words we can write any element in L as

$$\sum_{i,j} b_{ij}\alpha_i\beta_j$$

and so the mn elements $\alpha_i\beta_j$ span L as a vector space over F, now we must show they are linearly independent.

Suppose these elements are not independent, that is

$$\sum_{i,j} b_{ij}\alpha_i\beta_j = \sum_i a_i\alpha_i = 0$$

Since α_i form a basis of L over K, this means all the a_i must be zero

$$a_i = b_{i1}\beta_1 + \cdots + b_{in}\beta_n = 0$$

but since β_i are a basis for K over F then all the b_{ij} must also be zero. Thus the $\alpha_i\beta_j$ form a basis for L over F and

$$[L : F] = mn = [L : K][K : F]$$

as required. The second statement follows because if $[K : F]$ is infinite then there are infinitely many linearly independent elements of K (and thus of L) over F and so $[L : F]$ is also infinite. If $[L : K]$ then there are infinitely many linearly independent elements of L over K which are thus independent over F as well and so $[L : F]$ is infinite. The proof above shows that if $[L : K], [K : F]$ are finite then $[L : F]$ is finite so if $[L : F]$ is infinite then one of the two must also be infinite, thus proving the second statement. □

This relationship can be used in computations, for instance we have the corollary

Corollary 8.16. Suppose L/F is finite and let K be a subfield of L containing F ($F \subseteq K \subseteq L$), then

$$[K : F] \mid [L : F]$$

Example. Let $\sqrt[6]{2}$ be the 6th root of 2, then it's easy to see that

$$[\mathbb{Q}(\sqrt[6]{2}) : \mathbb{Q}] = 6$$

Since $(\sqrt[6]{2})^3 = \sqrt{2}$ we also have $\mathbb{Q}(\sqrt{2}) \subset \mathbb{Q}(\sqrt[6]{2})$ and we can compute

$$\overbrace{\underbrace{F \quad \subset}_{2} \mathbb{Q}(\sqrt{2}) \underbrace{\quad \subset \quad \mathbb{Q}(\sqrt[6]{2})}_{3}}^{6}$$

Definition. *A field extension K/F is finitely generated if $K = F(\alpha_1, \alpha_2, \ldots, \alpha_k)$*

Recall that a field generated over F is the smallest subfield of F which contains all the generators. It can be tedious to attempt to fit in all of the generators so we are lucky to have the following lemma.

Lemma 8.17. $F(\alpha, \beta) = (F(\alpha))(\beta)$, that is the field generated over F by α and β is the same as the field generated byu β over $F(\alpha)$

Proof. The field $F(\alpha, \beta)$ contains the field F and element α by definition and so contains $F(\alpha)$. By minimality we have

$$(F(\alpha)(\beta) \subseteq F(\alpha, \beta)$$

But $(F(\alpha)(\beta)$ contains F, α, β so by the minimality of $F(\alpha, \beta)$ we get the opposite inclusion

$$F(\alpha, \beta) \subseteq (F(\alpha)(\beta)$$

and so the two fields are equal. □

Thus we can compute the generated field through a sequence of simple extensions. In practice we compute a simple extension $F(\alpha)$ but appending it to the list of elements and attempting to fix all the issues they cause. Generally this involves appending the powers $\alpha^2, \alpha^3, \ldots$ in order the close the set under addition and multiplication.

Example. Consider the extension $\mathbb{Q}(\sqrt{2}, \sqrt{3})$. We already know that

$$[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$$

What remains is to find the degree of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ over $\mathbb{Q}(\sqrt{2})$. Since $\sqrt{3}$ is over degree 2 over $\mathbb{Q}$, this extension is of degree at most 2 and is precisely 2 if $x^2 - 3$ is irreducible over $\mathbb{Q}(\sqrt{2})$. Suppose it is reducible so that

$$\sqrt{3} \in \mathbb{Q}(\sqrt{2}) \to \sqrt{3} = a + b\sqrt{2} \to 3 = a^2 + 2b^2 + 2ab\sqrt{2}$$

Note that $ab = 0$ because otherwise this would imply $\sqrt{2}$ is rational, $b \neq 0$ because otherwise $\sqrt{3} = a$ is rational, and $a \neq 0$ because otherwise $\sqrt{6} = b$ is rational. Thus $ab = 0$ but yet $a, b \neq 0$ which is a contradiction and so $\sqrt{3} \notin \mathbb{Q}(\sqrt{2})$. Therefore we get

$$[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}] = [\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}(\sqrt{2})][\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2 \cdot 2 = 4$$

We can proceed by appending $\sqrt{3}$ to $\mathbb{Q}(\sqrt{2})$ and closing the field, resulting in

$$\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \left\{ a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6} \mid a, b, c, d \in \mathbb{Q} \right\}$$

We can now state the partial converse and thus characterize all finite extensions.

Theorem 8.18. An extension K/F is finite if and only if K is finitely generated by algebraic elements over F. In other words, a field generated over F by elements of degree $n_1, n_2, \ldots, n_k$ is algebraic of degree at most $n_1 n_2 \cdots n_k$.

Proof. Suppose K/F is finite of degree n and let $\alpha_1, \ldots, \alpha_n$ be a basis for K as a vector space over F. By Corollary 13.16, $[F(\alpha_i) : F]$ divides $[K : F]$ which means they are finite and thus the α_i algebraic. Therefore K is finitely generated by algebraic elements over F.

Conversely suppose K is finitely generated by algebraic elements. Then by Lemma 8.17 we can write

$$K = F(\alpha_1, \ldots, \alpha_k) = (F(\alpha_1, \ldots, \alpha_{k-1}))(\alpha_k)$$

this gives us a chain of fields

$$F = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_k = K$$

where $F_{i+1} = F_i(\alpha_{i+1})$. Suppose the generating elements have degree $n_1, \ldots, n_k$. Since each intermediate extension is simple, it must have degree at most n_i and so

$$[K : F] = [F_k : F_{k-1}] \cdots [F_2 : F_1][F_1 : F_0] \leq n_1 n_2 \cdots n_k$$

Thus the extension K/F is finite and this also proves the second statement. □

Corollary 8.19. Suppose α, β are both algebraic over F, then $\alpha \pm \beta, \alpha\beta, \alpha/\beta$ ($\beta \neq 0$ for the last one) are all algebraic over F.

Proof. These all lie in $F(\alpha, \beta)$ which is finite over F by the theorem, thus they are all algebraic by Corollary 13.14. □

This also proves

Corollary 8.20. Let L/F be an extension, the collection of elements in L that are algebraic over F form a subfield of L.

Example. Consider the extension $\mathbb{C}/\mathbb{Q}$ and let $\overline{\mathbb{Q}}$ be the subfield of all algebraic elements in $\mathbb{C}$ over $\mathbb{Q}$. Note that $\sqrt[n]{2} \in \overline{\mathbb{Q}}$ for all $n > 1$ which implies

$$[\overline{\mathbb{Q}} : \mathbb{Q}] \geq n$$

for all $n > 1$. Thus $\overline{\mathbb{Q}}$ is an infinite algebraic extension of $\mathbb{Q}$, called the algebraic numbers.

The property of a field being algebraic over another field is transitive.

Theorem 8.21. If K is algebraic over F and L is algebraic over K, then L is algebraic over F.

Proof. Consider some $\alpha \in L$ which is algebraic over K so it satisfies some polynomial

$$a_n\alpha^n + \cdots + a_1\alpha + a_0 = 0 \qquad a_i \in K$$

Consider the field $F(\alpha, a_0, \ldots, a_n)$ generated over F. Since K/F is algebraic, each of the elements a_i are algebraic over F. Thus the extension $F(a_0, \ldots, a_n)/F$ is finite by Theorem 8.17. We see that α generates an extension of degree at most n because it's minimal polynomial over F divides the one above (Corollary 13.11). Thus we compute

$$[F(\alpha, a_0, \ldots, a_n) : F] = [F(\alpha, a_0, \ldots, a_n) : F(a_0, \ldots, a_n)][f(a_0, \ldots, a_n) : F] < \infty$$

Thus $F(\alpha, a_0, \ldots, a_n)$ is finite and thus algebraic, in particular α is algebraic over F and therefore L is algebraic over F. □

A subfield generated by elements $\alpha_1, \ldots, \alpha_k$ over F where $\alpha_i \in K$ is the smallest subfield which contains the simple extensions $F(\alpha_i)$. We can extend this idea to subfields in general rather than just sets of elements.

Definition. *Let K_1, K_2 be subfields of K. The composite fields of K_1 and K_2, denoted K_1K_2 is the smallest subfield of K which contains both K_1 and K_2. This can also be extended to more than 2 subfields.*

The composite field can be thought of as the intersection of all subfields of K which contain both K_1 and K_2.

Example. The composite field $\mathbb{Q}(\sqrt{2})\mathbb{Q}(\sqrt[3]{2})$ is just $\mathbb{Q}(\sqrt[6]{2})$.

Proposition 8.22. Let K_1, K_2 be finite extensions of F contained in K, then

$$[K_1K_2 : F] \leq [K_2 : F][K_1 : F]$$

with equality only if a basis for one field over F remains linearly independent over the other field.

Proof. If $\alpha_1, \ldots, \alpha_n$ form a basis for K_1 over F and $\beta_1, \ldots, \beta_m$ form a basis for K_2 over F, then

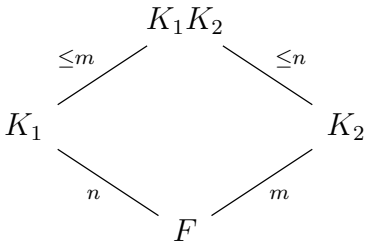
$$K_1K_2 = F(\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_m) = K_1(\beta_1, \ldots, \beta_m)$$

This implies that $[K_1K_2 : K_1] \leq m$ and since $[K_1K_2 : F] = [K_1K_2 : K_1][K_1 : F]$ this proves

$$[K_1K_2 : F] \leq mn = [K_2 : F][K_1 : F]$$

as required. □

Pictorially, this can be represented as



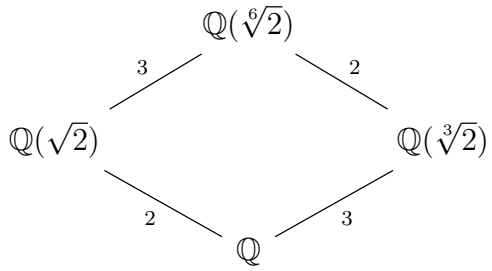
There's a simple situation in which equality can also hold.

Corollary 8.23. Suppose $[K_1 : F] = n$ and $[K_2 : F] = m$, then if $(n, m) = 1$ (that is they are relatively prime)

$$[K_1K_2 : F] = [K_2 : F][K_1 : F] = nm$$

Proof. The extension degree $[K_1K_2 : K]$ is divisible by n and m since K_1K_2 contain K_1 and K_2. Thus it is also divisible by their least common multiple, but if $(m, n) = 1$ then this is just mn. Since $[K_1K_2 : K] \leq mn$, this implies that the degree is mn. □

Example. From the corollary we see that $[\mathbb{Q}(\sqrt{2})\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 6$ and we get the diagram



8.3 Algebraic Closures

We proved in a previous section for for any polynomial $f(x) \in F[x]$ there exists a field extension K of F which contains a root α of $f(x)$. This is equivalent to saying that $f(x)$ has a factor $x - \alpha$ in the field K.

Definition. *An extension K/F is called a splitting field for $f(x) \in F[x]$ if $f(x)$ factors completely into linear factors (or splits completely) over K. Furthermore $f(x)$ does not factor over any proper subfield of K containing F. In this case we say that f splits in K.*

Theorem 8.24. Let F be a field and $f(x) \in F[x]$ a polynomial, then there exists an extension K/F which is a splitting field for $f(x)$.

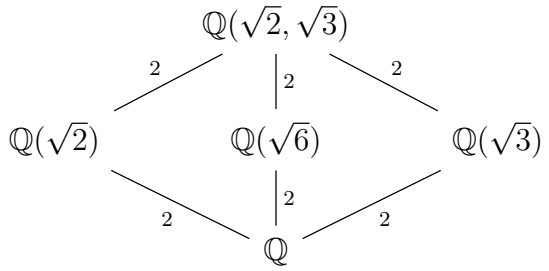
Proof. We can prove this through induction, let $f(x) \in F[x]$ be of degree n. If $n = 1$ then we F is already a splitting field for f. Otherwise suppose $n > 1$ and consider the irreducible factors of f. If all the irreducible factors are of degree 1, then F is already a splitting field and we are done. Otherwise there must be a factor $p(x)$ of degree at least 2, which means we can construct an extension E of F which contains a root of $p(x)$. The remaining factors have degree $n - 1$ and thus already have a splitting field by induction and so we are done. If K is the intersection of all the subfields of E which contain all the roots of $f(x)$, then K is the splitting field for $f(x)$. □

Example. The splitting field for $x^2 - 2$ is $\mathbb{Q}(\sqrt{2})$.

Definition. *If K is an algebraic extension of F which is the splitting field for some collection of polynomials $f(x) \in F[x]$, then K is called a normal extension of F.*

Generally we'll use the term splitting field instead of normal extension.

Example. The splitting field of $(x^2 - 2)(x^2 - 3)$ is $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ and we have the diagram



All splitting fields for a polynomial $f(x)$ are isomorphic and thus we can refer to *the* splitting field for a polynomial. This result comes from an extension of a previous theorem, Theorem 8.9.

Theorem 8.25. Let $\varphi : F \to F'$ be an isomorphism of fields and $f(x) \in F[x]$, $f'(x) \in F'[x]$ be polynomials where $f'(x)$ is obtained by applying φ to $f(x)$. If E, E' are splitting fields

for $f(x)$ and $f'(x)$ respectively, then φ extends to an isomorphism $\sigma : E \to E'$.

$$\begin{array}{ccc} E & \xrightarrow{\sigma} & E' \\ | & & | \\ F & \xrightarrow{\varphi} & F' \end{array}$$

Proof. We will prove this again using by induction on $deg(f) = n$. Recall that Theorem 8.9 proved that φ could be extended to an isomorphism $F[x] \cong F'[x]$ which implies that irreducible factors will map to irreducible factors. In the case $n = 1$ or when f only has irreducible factors of degree 1, linear factors gets mapped to linear factors so $f(x)$ splits over F and thus $f'(x)$ splits over F' (linear factors are also preserved by φ). Thus $E = F \cong F' = E'$ and we can take $\sigma = \varphi$ as the extension.

Suppose this holds for all degrees less than n and let $p(x)$ be an irreducible factor of degree at least 2 which corresponding factor $p'(x)$. If $\alpha \in E$ is a root of $p(x)$ and $\beta \in E'$ a root of $p'(x)$, we can construct an isomorphism $\sigma' : F(\alpha) \to F'(\beta)$ leading to the diagram.

$$\begin{array}{ccc} F(\alpha) & \xrightarrow{\sigma'} & F'(\beta) \\ | & & | \\ F & \xrightarrow{\varphi} & F' \end{array}$$

Thus over $F(\alpha)$, f splits as $f(x) = (x - \alpha)f_1(x)$ and f' splits over $F'(\beta)$ as $f'(x) = (x - \beta)f_1'(x)$. The remaining factor f_1, f_1' have degree $n - 1$. The field E is a splitting field for $f_1(x)$ over $F(\alpha)$ because it contains all the roots of $f_1(x)$ and if any smaller extension L contains F_1 and these roots, then it would contain all the roots of $f(x)$ and be a smaller splitting field, a contradiction. A similar argument shows that E' is a splitting field for $f_1'(x)$ over $F(\beta)$. Furthermore induction provides another extension $\sigma : E \to E'$ of σ', giving the diagram

$$\begin{array}{ccc} E & \xrightarrow{\sigma} & E' \\ | & & | \\ F(\alpha) & \xrightarrow{\sigma'} & F'(\beta) \\ | & & | \\ F & \xrightarrow{\varphi} & F' \end{array}$$

The diagram also shows the desired restrictions which shows that σ is an extension of φ, completing the proof. □

Corollary 8.26. Any two splitting fields for a polynomial $f(x) \in F[x]$ over a field F are isomorphic.

Proof. Let $\varphi : F \to F$ be the identity map and E, E' be two splitting fields for $f(x)$, then the previous theorem implies $E \cong E'$. □

We will now discuss field extensions which contain the roots of every polynomial over a field F. These are significant because oftentimes we want to consider a family of polynomials over a certain field and it's pointless to try to construct a field extension for every individual polynomial.

Definition. *The field $\overline{F}$ is an algebraic closure of F if it is algebraic over F and every polynomial $f(x) \in F[x]$ splits completely over $\overline{F}$.*

In other words, $\overline{F}$ contains all the elements algebraic over F.

Definition. *A field K is algebraically closed if every polynomial with coefficients in K has a root in K.*

The names "algebraic closure" and "algebraically closed" seem to imply some sort of correlation, lets examine this further.

Proposition 8.27. An algebraic closure $\overline{F}$ of a field F is algebraically closed.

Proof. Consider some polynomial $f(x) \in \overline{F}[x]$ and suppose α is a root of $f(x)$. This root generates an algebraic extension $\overline{F}(\alpha)$ of $\overline{F}$. Since$\overline{F}$ is algebraic over F, by Theorem 8.21 this implies that $\overline{F}(\alpha)$ is algebraic over F and that α is algebraic over F. But $\alpha \in \overline{F}$ so $\overline{F}$ is algebraically closed. □

Since we can construct splitting fields for any polynomial we may also ask if it is possible to construct an algebraic closure for any field F, the answer is yes. We start with the proposition...

Proposition 8.28. For any field F, there exists an algebraically closed field K containing F.

Proposition 8.29. Let K be an algebraically closed field containing a subfield F. Then the collection of elements

$$\overline{F} = \{\alpha \in K \mid \alpha \text{ is algebraic over } F\}$$

forms an algebraic closure over F.

Proof. $\overline{F}$ is an algebraic extension by definition, each polynomial splits into linear factors $x - \alpha$ over K. Since each α is the root of some polynomial, they are algebraic and thus $\alpha \in \overline{F}$. Thus all the linear factors have coefficients in $\overline{F}$ and any polynomial in $F[x]$ will split completely in $\overline{F}[x]$. Therefore $\overline{F}$ is an algebraic closure of F. □

These two propositions combined prove the existence of algebraic closures.

Corollary 8.30. Every field has an algebraic closure

Proof. For any field F we can construct an algebraically closed field K containing F. The collection of elements in K which are algebraic over F form an algebraic closure. □

A useful result that can be proved using Galois theory is the Fundamental Theorem of Algebra which we will state in terms of algebraic closures.

Theorem 8.31 (Fundamental Theorem of Algebra)**.** The field $\mathbb{C}$ is algebraically closed.

The more standard statement is that every polynomial in $\mathbb{C}[x]$ will have at least one complex root. This agrees with our version by definition.

9 Galois Theory

9.1 Splitting Fields

We introduced splitting fields in the previous section, but let's discuss them a bit more. Recall that the splitting field is the smallest extension in which a polynomial splits.

Example. Let p be prime and take $f = x^{p-1} + x^{p-2} + \cdots + 1$. Note that our polynomial is just

$$f(x) = \frac{x^p - 1}{x - 1}$$

The roots of $x^p - 1$ are just the pth roots of unity (but not 1) and since p is prime we can get any other root of unity through exponentiation. Thus if we add any root to $\mathbb{Q}$ then f will split completely, so the splitting field is $\mathbb{Q}[\xi]$

Example. As a similar example consider $f(x) = x^p - x - a$ which is irreducible over $\mathbb{F}_p$. If we add some root α, then we get $\alpha^p - \alpha = a$ and the remaining roots are $\alpha+1, \alpha+2, \ldots, \alpha+p-1$.

$$(\alpha + 1)^p - (\alpha + 1) = \alpha^p + 1 - \alpha - 1 = a$$

So the splitting field is just $L = \mathbb{Q}[\alpha]$. In any extension where f has a root it will also have p roots, that is $[L : \mathbb{F}_p] = p$.

Definition. *A polynomial $f \in k[x]$ has simple roots if it splits in some extension as*

$$f = \prod_{i=1}^{deg\ f} (x - \alpha_j)$$

where $\alpha_i \neq \alpha_j$ for $i \neq j$.

In other words a polynomial has simple roots if it splits into linear terms with distinct roots.

Proposition 9.1. If k is a field and $f \in k[x]$, let F be generated by roots of f and suppose f splits in L, then

$$[F : k] \geq \{\text{number of } k\text{-homs } F \to L\} \geq 1$$

If we have equality, then f only has simple roots.

We will prove this later, first we will examine more closely what it means to have multiple roots.

Lemma 9.2. f has no multiple roots if and only if

$$gcd(f, f') = 1$$

where f' is the standard derivative

$$\frac{d}{dx}(ax^n) = anx^{n-1}$$

Proof. If f has multiple roots, then it is of the form

$$f = (x - a)^2 \cdot g(x)$$

which gives a derivative

$$f' = 2(x - a) \cdot g(x) + (x - a)^2 \cdot g'(x)$$

and thus the greatest common divisor is not one. □

Example. Suppose k is of characteristic p and consider $f = x^p - a$. The derivative is 0 so f has multiple roots, in fact all of its roots are identical.

$$x^p - a = (x - \alpha)^p$$

What happened in the example is pretty common in fields of characteristic p, in fact we even have a lemma for it.

Lemma 9.3. If $f \in k[x]$ is irreducible, then the following are equivalent:

1. f has multiple roots
2. char $k = p$ and $f = g(x^p)$
3. all roots of f are multiple

Proof. $1 \to 2$: If f has multiple roots, then

$$gcd(f, f') \neq 1$$

but note then that

$$\deg f' < \deg f \to \deg gcd(f, f') < \deg f$$

unless $f' = 0$. If it's nonzero, then

$$gcd(f, f') \mid f$$

which is a contradiction. Thus we can write

$$f = \sum_{i=0} a_i x^i \qquad f' = \sum_{i=1} i a_i x^{i-1}$$

Since f' has to be zero, then $a_i = 0$ for all $p \nmid i$, in other words

$$f = g(x^p)$$

$2 \to 3$: Assume that we have

$$f = a_0 x^{np} + a_1 x^{p(n-1} + \cdots + a_n$$

Take some extension field k in which $a_0, \ldots, a_n$ have pth roots $b_0, \ldots, b_b$. Then in k we have

$$f = (b_0 x^n + \cdots + b_n)^p = g^p$$

If g further splits then all the roots of f are multiple.

$3 \to 1$: obvious □

Definition. *A polynomial is separable if it is nonzero and has distinct (simple) roots.*

Example. $x^p - a$ is never separable since it is irreducible in characteristic 0 and has multiple roots in characteristic p.

Definition. *A field k is perfect if it is either characteristic 0 or is characteristic p and every $a \in k$ admits a pth root.*

The motivation behind perfect fields is that if we work in one, then every irreducible polynomial will have simple roots. Thus every irreducible polynomial in a perfect field is separable.

Example. Some examples of perfect fields

- finite fields
- algebraically closed fields
- fields of characteristic 0

Lemma 9.4. Let $F = k[\alpha]$ where α is algebraic over K, define $f = min(\alpha)$ and suppose it splits in G/k. Then there exists a map $\phi : F \to G$ over k with

$$\text{number of such } \phi \leq [F : k]$$

with equality if f has simple roots in G.

Proof. There is a one-to-one correspondance

$$\{\phi\} \iff \{\text{rotos of } f \text{ in } G\}$$

Thus we have

$$\text{number of } \phi = \text{distinct roots of } f \leq \deg f = [F : k]$$

□

Now we prove the general case which we presented earlier.

Proposition 9.5. Let $f \in k[x]$ be a polynomial. Suppose we have a "small" extension F/k which means $F = k[\alpha_1, \ldots, \alpha_n]$ where α_i are the roots of f and a "big" extension G/k which means f splits in G. Then there exists a k-homomorphism $\phi : F \to G$, furthermore

$$\text{number of such } \phi \leq [F : k]$$

with equality if f has simple roots.

Proof. We can simply repeat the process of adjoining roots. Define $L_1 = k[\alpha_1] \subseteq F$, by the lemma we have

$$1 \leq \text{number of } \phi : L_1 \to G \leq [L_1 : k]$$

Now we take α_2, which is algebraic over k and thus L_1. Let $g = min_{L_1}(\alpha_2)$ which divides f in L_1 so g will have distinct roots in f also does. Note that $L_2 = L_1[\alpha_2] = k[\alpha_1, \alpha_2]$. Thus by the lemma

$$\begin{aligned} 1 \leq \text{number of } L_1\text{-homomorphisms } \phi : L_2 \to G &\leq [L_2 : L_1] \\ 1 \leq \text{number of } k\text{-homomorphisms } \phi : L_2 \to G &\leq [L_2 : L_1][L_1 : k] \\ &\leq [L_2 : k] \end{aligned}$$

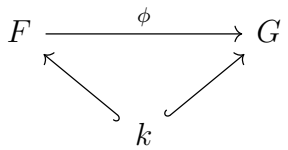
Repeating this process for all roots gives the desired inequality. □

If F, G are both splitting fields (and so are both "big" and "small"), then all such ϕ are isomorphisms and there are $[F : k]$ of them.

9.2 The Fundamental Theorem of Galois Theory

The fundamental theorem of Galois theory gives a connection between the field extensions we've been studying and groups. The group elements are automorphisms which we will study first before presenting the theorem.

Definition. *Let k be a field with extensions F and G. A k-homomorphism (or homomorphism over k) is a ring map $\phi : F \to G$ such that the diagram commutes*



A k-isomorphism is also called a k-automorphism.

In other words it's just a ring map such that the underlying field is kept the same. The set of invertible k-homomorphisms forms the automorphism group, denoted

$$Aut(L/K) = \{\phi : L \to L \mid \phi \text{ an invertible } k\text{-homomorphism}\}$$

Example. The automorphism group can be trivial even thought the extension isn't, for instance

$$Aut(\mathbb{Q}[\sqrt[3]{2}]/\mathbb{Q}) = \{1\}$$

An extension having no automorphisms is not something desirable, in this case we may think of $\mathbb{Q}[\sqrt[3]{2}]$ as a "bad" extension. We can also see that it's not a "good" extension because it is not a splitting field. There is a useful correspondence between automorphisms and field elements.

Theorem 9.6. Let $F = k(\alpha)$ and G/k be two field extensions, then

1. If α is transcendental

$$\{\phi : F \to G \text{ over } k\} \iff \{\beta \in G \mid \beta \text{ transcendental over } k\}$$

2. If α is algebraic and $f = min(\alpha)$ its minimal polynomial

$$\{\phi : F \to G \text{ over } k\} \iff \{\beta \in G \mid f(\beta) = 0\}$$

where both correspondences are one to one.

Proof. The correspondence is given by

$$\{\phi\} \iff \{\phi(\alpha)\}$$

□

The automorphism group can also be represented as

$$Aut(E/F) = \{\phi : E \to E \mid \phi \text{ an automorphism, } \phi|_F = id\}$$

Definition. *Let $G \leq Aut(E)$ be a subgroup, define*

$$E^G = \{x \in E \mid g \cdot x = x \ \forall g \in G\}$$

This is called the subfield of G-invariants of E or the fixed field of G,

Example. The automorphism group of $\mathbb{C}/\mathbb{R}$ consists of two elememts: the identity and conjugation. Thus

$$G = Aut(\mathbb{C}/\mathbb{R}) \cong \mathbb{Z}/2\mathbb{Z}$$

The fixed field $\mathbb{C}^G$ is just all the real numbers.

The situation above in which the fixed field is the underlying field is highly preferable. Note that in general

$$F \subseteq E^{Aut(E/F)}$$

but we do not want the inclusion to be strict (so we want equality instead).

Example. Take $F = \mathbb{Q}$ and $E = \mathbb{Q}[\sqrt[3]{2}] \subseteq \mathbb{R}$. The only root of $x^3 - 2$ in E is the real root, so

$$Aut(E/F) = \{id\}$$
$$E^{Aut(E/F)} = E \subset F$$

which is not desirable. We see that this occurs because $x^3 - 2$ does completely split in E.

Example. Suppose F is of characteristic p and $\alpha \in F$ does not have a pth root (F isn't perfect). Let E be the splitting field for $x^p - \alpha$, in E we have

$$x^p - \alpha = (x - \beta)^p$$

where β is some root. But even though $x^p - \alpha$ splits completely, it still only has one root, thus

$$Aut(E/F) = \{id\}$$
$$E^{Aut(E/F)} = E \subset F$$

Now the issue isn't that $x^p - \alpha$ doesn't completely split, it is that the roots aren't distinct so that f is separable.

There is an easy way to count the number of automorphisms for an extension given that it is a splitting field.

Theorem 9.7. If E/F is the splitting field of a separable polynomial over F then

$$|Aut(E/F)| = [E : F]$$

Proof. Let E/F be the splitting field of some $f \in F[x]$, then it contains the deg f distinct roots of f. By lemma 9.4 we know that the number of F-homomorphisms $E \to E$ is $[E : F]$. Since E is finite over F, all such homomorphisms are isomomorphisms (automorphisms). □

Theorem 9.8 (Artin)**.** Let $G \leq Aut(E)$ be finite, then

$$[E : E^G] \leq |G|$$

Proof. Let $G = \{\sigma_1, \ldots, \sigma_m\}$ with $m = |G|$ and labeled such that $\sigma_1 = id$. Thus if $[E : E^G] \leq |G|$ then every collection of $n > m$ elements in E is linearly dependent over E^G (recall that $[E : E^G]$ is the dimension of E over E^G).

Let's start with a collection $\alpha_1, \ldots, \alpha_n \in E$. We want to show that this collection is linearly dependent, that is there exists $c_i \in F$ such that

$$c_1\alpha_1 + \cdots + c_n\alpha_n = 0$$

Since $\sigma_i \in Aut(E/F)$ (F is fixed by σ_i), if the above equality holds then for all i

$$c_1\sigma(\alpha_i) + \cdots + c_n\sigma_i(\alpha_n) = 0$$

using this fact we can construct a system of m equations in E to find the c_i

$$\sigma_1(\alpha_1)x_1 + \cdots + \sigma_1(\alpha_n)x_n = 0$$
$$\vdots$$
$$\sigma_m(\alpha_1)x_1 + \cdots + \sigma_m(\alpha_n)x_n = 0$$

Note that this system is homogeneous, thus there must exist some solution $(c_1, \ldots, c_n)$ in which the c_i are not all zero. Let's choose one with the largest number of zeros, we claim that this solution is in F^n, that is they all lie in F.

Start by reordering the c_i such that $c_1 \neq 0$ and rescaling (e.g. by dividing by c_1) so that $c_1 \in F$. Assume that there exists some $c_i \neq 0 \notin F$ (since otherwise we are done), then there exists some k such that $\sigma_k(c_i) \neq c_i$ (since c_i is no longer fixed as it's not in F). But since σ_i fixes F we can just reapply σ_k to our entire system of equations. Because G is finite, σ_k simply reorders the lines, showing that we have another solution

$$(\sigma_k(c_1), \ldots, \sigma_k(c_n))$$

Now since we've reordered the solutions such that $c_i \in F$, this solution is just

$$(c_1, \sigma_k(c_2), \ldots, \sigma_k(c_n)$$

which proves, through subtracting the original solution, that we have another solution

$$(0, c_2 - \sigma_k(c_2), \ldots, c_i - \sigma_k(c_i), \ldots, c_n - \sigma_k(c_n))$$

Now note that if $c_j = 0$ then $c_j - \sigma_k(c_j) = 0$, furthermore $c_i \neq \sigma_k(c_i)$ implies $c_i - \sigma_k(c_i) \neq 0$. Thus this solution is nonzero (ith term is nonzero) and has one more zero (0's are preserved and first term is now 0), a contradiction. □

Corollary 9.9. Let $G \leq Aut(E)$ be finite, then

$$G = Aut(E/E^G)$$

Proof. Since $G \leq Aut(E)$ is a subgroup, we have the following inequalities (using the previous theorem)

$$[E : E^G] \leq |G| \leq |Aut(E/E^G)| \leq [E : E^G]$$
$$\therefore G = Aut(E/E^G)$$

□

Now that we've sufficiently studied automorphisms it's time to define what we mean by a "good" extension.

Definition. *Let E/F be an extension of fields*

1. *An element $x \in E$ is separable if it's minimal polynomial is separable (i.e. has all distinct roots in some extension).*
2. *The extension E/F is separable if every element in E is separable over F.*
3. *E/F is normal if the for all $x \in E$, its minimal polynomial over F splits in E.*
4. *E/F is Galois if it is finite, separable, and normal.*

First let's recall some field extensions we studied earlier which are not Galois.

Example. $\mathbb{Q}[\sqrt[3]{2}]/\mathbb{Q}$ is not Galois. If is finite (of degree 3), separable (since characteristic 0), but not normal because $min_{\mathbb{Q}}(\sqrt[3]{2}) = x^3 - 2$ does not split.

Example. $\mathbb{F}_p(x)/\mathbb{F}_p(x^p)$ is normal but not separable because the minimal polynomial of x is not separable.

Now we present some alternative characterizations of Galois extensions.

Theorem 9.10. For an extension E/F, the following are equivalent

1. E is the splitting field of some separable $f \in F[x]$
2. E is finite over F and $F = E^{Aut(E/F)}$
3. $F = E^G$ for some finite $G \leq Aut(E)$
4. E is Galois over F

Proof. $1 \to 2$: If E is a splitting field over F, then it is certainly finite. Define

$$F' = E^{Aut(E/F)} \supseteq F$$

we want to show that $F' = F$. Note that E can also be regarded as a splitting field for f over F' and that f is still separable when treated this way. Thus

$$|Aut(E/F')| = [E : F'] \leq [E : F] = |Aut(E/F)|$$

But $|Aut(E/F')| = |Aut(E/F)|$ (Corollary 9.9), so $[E : F'] = [E : F]$ and $F = F'$.

$2 \to 3$: Let $G = Aut(E/F)$, by assumption we have $F = E^G$ which means G is finite since E is finite over F.

$3 \to 4$: From theorem 9.8 we have $[E : F] < |G|$, in particular E/F is finite. Now we need to show that it is separable and normal. Consider some $\alpha \in E$ and let $f = min(\alpha)$. Let $\{\alpha_1 = \alpha, \alpha_2, \ldots, \alpha_n\}$ be the orbit of α under the action of G (note the α_i are distinct). Define the polynomial

$$g(x) = \prod_{i=1}^{n}(x - \alpha_i) = x^n + a_1 x^{n-1} + \cdots + a_n$$

The coefficients a_j are given by symmetric polynomials (by binomial expansion) in α_i, thus each $\sigma \in G$ simply permutes the α_i which preserves the a_j. In other words $\sigma a_j = a_j$ for all j which means $g \in F[x]$ as the coefficients are fixed by G.

Since $g(\alpha) = 0$, it must be divisible by f. Let $\alpha_i = \sigma\alpha$, when we apply σ to the equation $f(\alpha) = 0$ we get $f(\alpha_i) = 0$. Therefore every α_i is a root of f and so g divides f. Thus $f = g$ and we see that f splits into distinct factors in E. This implies that the minimal polynomials for all elements in E are separable and split, thus E is Galois.

$4 \to 1$: Since E is finite over F, it must be of the form $E = F[\alpha_1, \ldots, \alpha_n]$ where $\alpha_i \in E$ are algebraic over F. Let $f_i = min(\alpha_i)$ be the minimal polynomials and let f be the product of all the f_i. Because E is normal over F, each f_i splits in E and so E is the splitting field over f. Furthermore since E is separable over F, each f_i is separable is thus so is f. □

If we have some group finite group G of automorphisms over a field E, then we can trivially construct a Galois extension by considering E/E^G.

Corollary 9.11. If E/F is Galois with $E \subset M \subset F$, then E/M is Galois but M/F may not be.

Proof. The same polynomial for which E is a splitting field over F can be considered as a polynomial in $M[x]$. □

Example. As a counterexample to the corollary's second statement, note

$$\underbrace{\mathbb{Q}[\sqrt[3]{2}, e^{2\pi i/3}] \quad \subseteq}_{\text{Galois}} \mathbb{Q}[\sqrt[3]{2}] \underbrace{\quad \subseteq \quad \mathbb{Q}}_{\text{Not Galois}}$$

Corollary 9.12. Every finite, separable extension is contained in a Galois extension.

Proof. Consider an extension E/F and let $E = F[\alpha_1, \ldots, \alpha_n]$ and $f_i = min(\alpha_i) \in F[x]$. The product of distinct f_i's will give a separable polynomial in $F[x]$ whose splitting field is a Galois extension of F which contains E. □

The automorphism group of a Galois extension is special, in fact there is a connection between the group structure of that group and the field structure of the extension.

Definition. *If E/F is Galois, then the group*

$$Gal(E/F) = Aut(E/F)$$

is called the Galois group.

Definition. *Let E/F be an extension of fields, a subextension is an extension M/F with $M \subseteq E$. Note that we thus have the inclusions $F \subseteq M \subseteq E$.*

Theorem 9.13 (Fundamental Theorem of Galois Theory)**.** Let E/F be a Galois extension with Galois group $G = Gal(E/F)$, then there is a one to one correspondence

$$\{\text{subgroups } H \leq G\} \iff \{\text{subextensions } E \supseteq M \supseteq F\}$$

given by

$$H \mapsto E^H$$
$$Gal(E/M) \hookleftarrow M$$

with the following properties:

1. The correspondence reverses inclusions
$$K \leq H \to E^k \supseteq E^H$$
2. The index of the subgroup is equal to the degree of the extension
$$|H_1 : H_2| = [E^{H_1} : E^{H_2}]$$
3. If $\sigma \in G$ and $H \leq G$ with $H \leftrightarrow M$, then
$$\sigma H \sigma^{-1} \leftrightarrow \sigma M$$
4. If $H \trianglelefteq G$ is normal, then E^H/F is normal (hence Galois) with
$$Gal(E^H/F) \cong G/H$$

Proof. First we must show that the given maps are inverses. Let $H \leq G$ be a subgroup, we have from corollary 9.9 that $Gal(E/E^H) = H$. Let M/F be a subextension, then E is Galois over M by corollary 9.11 which means $E^{Aut(E/M)} = M$. So the two maps are indeed inverses, now we prove the properties of these maps.

1) Note that
$$K \leq H \to E^K \supseteq E^H \to Gal(E/E^K) \subseteq Gal(E/E^H) \to K \leq H$$

2) Let $H_1 \leq G$ be a subgroup and suppose $H_2 = 1$ for now. Then
$$|Gal(E/E^{H_1}) : 1| = |Gal(E/E^{H_1})| = [E : E^{H_1}]$$
For some general subgroup H_2, the result follows from using
$$|H_1| = |H_1 : H_2||H_2|$$
$$[E : E^{H_1}] = [E : E^{H_2}][E^{H_2} : E^{H_1}]$$
to get
$$|H_1 : H_2| = \frac{|H_1|}{|H_2|} = \frac{[E : E^{H_1}]}{[E : E^{H_2}]} = [E^{H_2} : E^{H_1}]$$

3) For $\tau \in G$ and $\alpha \in E$
$$\tau\alpha = \alpha \iff \sigma\tau\sigma^{-1}(\sigma\alpha) = \sigma\alpha$$
In other words τ fixes M if and only if $\sigma\tau\sigma^{-1}$ fixes σM which would imply
$$\sigma Gal(E/M)\sigma^{-1} = Gal(E/\sigma M) \iff \sigma H \sigma^{-1} \leftrightarrow \sigma M$$

4) Let $H \trianglelefteq G$ be normal which implies $\sigma H \sigma^{-1} = H$ for all $\sigma \in G$. This suggests that the action of G on E stabilizes E^H since $\sigma E^H = E^H$ for all σ, giving a homomorphisms
$$\sigma \mapsto \sigma|_{E^H} : G \to Aut(E^H/F)$$

with a kernel of H. Since $(E^H)^{G/H} = F$, we see that E^H is Galois over F with

$$Gal(E^H/F) \cong G/H$$

Conversely suppose M is normal over F and is generated by $\alpha_1, \ldots, \alpha_n$ over F. $\sigma\alpha_i$ is a root of the minimal polynomial over α_i for all $\sigma \in G$ and thus lies in M. Thus $\sigma M = M$ which means $\sigma H \sigma^{-1} = H$ and H is normal. $\square$

Example. Consider the extension $E/F = \mathbb{Q}[\sqrt[3]{2}, e^{2\pi i/3}]/\mathbb{Q}$, which is of degree 6. Note that

$$|G : 1| = |G| = [E : F] = 6$$

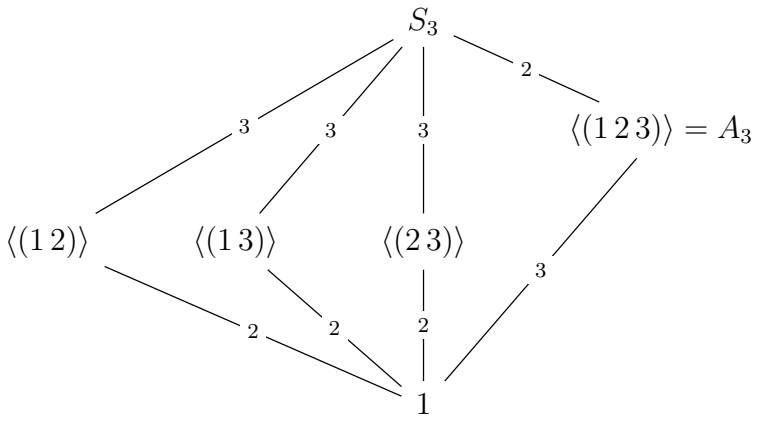
which gives two possibilities for the Galois group G

$$G = \begin{cases} \mathbb{Z}/6\mathbb{Z} & G \text{ is cyclic} \\ S_3 & G \text{ is not cyclic} \end{cases}$$

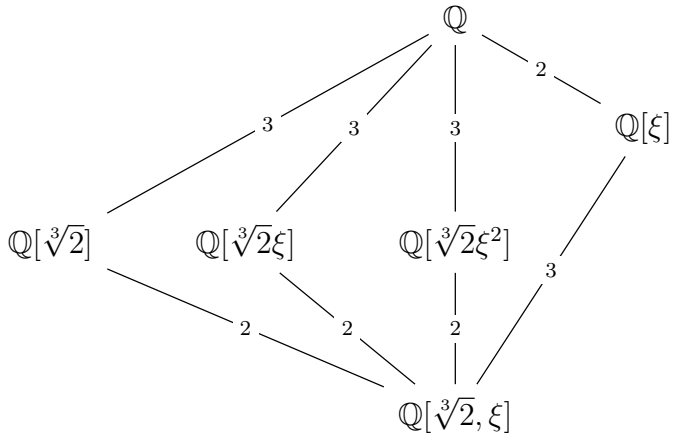
If G were cyclic, then all subgroups are normal. This would imply that all intermediate extensions are Galois which is clearly false because $\mathbb{Q}[\sqrt[3]{2}]$ is not Galois. Thus

$$G = Gal(E/F) = S_3$$

To calculate the intermediate subfields, first we note the existing subgroups



A_3 is particularly important since it is normal in A_3. The corresponding subfields will be upside down since the theorem is inclusion reversing.



The importance of A_3 is seen in this diagram, the associated field extension $\mathbb{Q}[\xi]$ is Galois whereas none of the other intermediate extensions are. However note that all the non-normal subgroups are conjugate, for instance

$$\begin{aligned}(2\,3)(1\,2)(2\,3)^{-1} &= (2\,3)(1\,2)(2\,3)\\ &= (2\,3)(1\,2\,3)\\ &= (1\,3)\\ \therefore (2\,3)\langle(1\,2)\rangle(2\,3)^{-1} &= \langle(1\,3)\rangle\end{aligned}$$

Thus from the theorem there must exists some $\sigma \in Gal(E/F)$ such that

$$\sigma\mathbb{Q}[\sqrt[3]{2}] = \mathbb{Q}[\sqrt[3]{2}\xi]$$

We can find σ using the fact that it is completely determined by where it sends $\sqrt[3]{2}$ and ξ. In our case we want

$$\begin{gathered}\sqrt[3]{2} \mapsto \sqrt[3]{2}\xi\\ \therefore \sqrt[3]{2}\xi \mapsto \sqrt[3]{2} \text{ since } \sigma^2 = id\\ \therefore \xi \mapsto \xi^2\end{gathered}$$

Example. Consider the extension $E/F = \mathbb{Q}[\xi]/\mathbb{Q}$ where $\xi = e^{2\pi i/7}$, the 7th root of unity. This is Galois because ξ is a root of

$$x^7 - 1 = (x-1)(x^6 + x^5 + \cdots + 1)$$

in particular it is the root of the irreducible second factor. The roots of this part are $\xi, \xi^2, \ldots, \xi^6 \in \mathbb{Q}[\xi]$ so this is the splitting field of this factor. An extension of this form is called a cyclotomic extension.

If $G = Aut(E/F)$ is the Galois group, then $|G| = 6$ and the automorphisms of E/F are characterized by

$$\begin{gathered}\xi \mapsto \xi^i\\ \mathbb{Q} \mapsto \mathbb{Q}\end{gathered}$$

This gives rise to a one to one correspondence

$$\{\phi : F[\xi] \to E \text{ over } F\} \iff \{\text{roots of } min_E(\xi) \text{ in } E\}$$

The six automorphisms given by this correspondence are all the elements of G, meaning that there is a mapping

$$G \to (\mathbb{Z}/7\mathbb{Z})^\times \qquad \left\{\sigma(\xi) = \xi^i\right\} \mapsto i$$

But $(\mathbb{Z}/7\mathbb{Z})^\times$ is cyclic which means we should be looking for a generator for G. First note that the powers of 3 generate $(\mathbb{Z}/7\mathbb{Z})^\times$

$$\left\{3^i \mod 7\right\} = \{3, 2, 6, 4, 5, 1\} = (\mathbb{Z}/7\mathbb{Z})^\times$$

If we pull back this action we can let

$$\sigma = (\xi \mapsto \xi^3) \qquad \langle\sigma\rangle = G$$

So now what are the subfields? Abstractly we have two subgroups of $G \cong \mathbb{Z}/6\mathbb{Z}$, these are

$$\langle 2\rangle \cong \mathbb{Z}/2\mathbb{Z} \to \langle\sigma^2\rangle\langle 3\rangle \cong \mathbb{Z}/3\mathbb{Z} \to \langle\sigma^3\rangle$$

Since G are abelian, these are both normal which means the associated subextensions will be Galois. To identify the subfield, we must find the subfield fixed by $\langle\sigma^3\rangle$

$$\mathbb{Q}[\xi]^{\langle\sigma^3\rangle} = \mathbb{Q}[\xi]^{\sigma^3} \to \sigma\xi = \xi^3$$

By doing some calculations to find how σ^3 acts, we find

$$\xi \mapsto \xi^3 \mapsto \xi^9\,(=\xi^2) \mapsto \xi^{27}\,(=\xi^6) = \overline{\xi}$$

So while ξ isn't invariant under σ^3, $\xi + \overline{\xi}$ is invariant. Thus the Galois subextension is

$$\mathbb{Q}[\xi + \overline{\xi}] = \mathbb{Q}[2\cos\frac{2\pi}{7}]$$

The remaining subextension is generated by $\beta = \xi + \xi^2 + \xi^4$. If we let $\beta' = \sigma\beta$, then we see that $(\beta - \beta')^2 = -7$. Thus the fixed field is $\mathbb{Q}[\sqrt{-7}]$ and we get the diagrams