Abstract Algebra

Muchang Bahng

Spring 2024

Contents

With set theory, we have established what sets, along with functions and relations are. Abstract algebra extends on this by studying *algebraic structures*, which are sets S with specific *operations* acting on their elements. This is a very natural extension and to be honest does not require much motivation. Let's precisely define what operations are.

Definition 0.1 (Operation)

A **p-ary operation**[a] $*$ on a set A is a map

$$* : A^p \longrightarrow A \tag{1}$$

where A^p is the p-fold Cartesian product of A. In specific cases,
1. If $p = 1$, then $*$ is said to be **unary**.
2. If $p = 2$, then $*$ is **binary**.

We can consider for $p > 2$ and even if p is infinite.

[a] or called an operation of arity p.

Definition 0.2 (Algebraic Structure)

An **algebraic structure** is a nonempty set A with a finite set of operations $*_1, \dots, *_n$ and satisfying a finite set of axioms. It is written as $(A, *_1, \dots, *_n)$.

If we consider functions between algebraic structures $f : A \to B$, there are some natural properties that we would like f to have.

Definition 0.3 (Preservation of Operation)

Given algebraic structures (A, μ_A), (B, μ_B), where μ_A and μ_B have the same arity p, a function $f : A \to B$ is said to **preserve the operation** if for all $x_1, \dots, x_p \in A$,

$$f(\mu_A(x_1, \dots, x_p)) = \mu_B(f(x_1), f(x_2), \dots, f(x_p)) \tag{2}$$

Functions that preserve operations are generally called *homomorphisms*. However, given that preservation is defined with respect to each operation, a map may preserve one operation but not the other. Therefore, we will formally define homomorphisms for each class of algebraic structures we encounter.

Definition 0.4 (Commutative, Associative Operations)

A binary operation $\cdot : A \times A \to A$ is said to be
1. **associative** if for all $a, b, c \in A$, $(ab)c = a(bc)$.
2. **commutative** if for all $a, b \in A$, $ab = ba$.

Associativity is a particularly important property that we would like to have, and it is quite rare to work with algebraic structures that don't have associativity. It basically states that when doing an operation sequentially over 3 elements, it doesn't matter if we evaluate ab or bc first. Therefore, associativity allows us to throw the parentheses away since the evaluated result does not change.

Commutativity on the other hand is not as prevalent. It simply tells us that we can "swap" terms when evaluating. This usually is a another nice convenience, and in the theory of rings commutativity is very prevalent. Either way, in both of these scenarios we can extend to any finite sequence of operations.

Theorem 0.1 (Generalized Associativity)

Given that a binary operation $\cdot$ is associative on a set S, it is always the case that for any finite collection $a_1, \ldots, a_n$, the value $a_1 \ldots a_n$ is unique.

Proof. We prove by strong induction on n from $n = 3$. Clearly $(a_1a_2)a_3 = a_1(a_2a_3)$ by definition of associativity. The rest is a bit tedious but is mentioned in Jacobson's *Basic Algebra 1*.

Theorem 0.2 (Generalized Commutativity)

Given that a binary operation $\cdot$ is commutative and associative on a set S, with $\alpha = a_1 + \ldots + a_n$, we have

$$\alpha = a_{i_1} + \ldots + a_{i_n} \tag{3}$$

for any permutation $(i_1, \ldots, i_n)$ of $(1, \ldots, n)$.

Now that we've gotten these out of the way, we can start talking about algebraic structures. I've went through 4 main textbooks, plus Google and talking to friends/professors in creating these notes.

1. Vinberg's *A Course in Algebra*.
2. Nathan Jacobson's *Basic Algebra 1*, given to me by Marty.
3. Ted Shifrin's *Abstract Algebra, A Geometric Approach*, used in Duke Math 401.
4. Dummit and Foote's *Abstract Algebra, 3rd Edition*, used in Duke Math 501.

1 Group-Like Structures

1.1 Semigroups and Monoids

Now the endowment of some structures gives rise to the following. Usually, we will start with the most general algebraic structures and then as we endow them with more structure, we can prove more properties. Let's talk about the most basic type of algebraic structure. If you have a set S and some associative operation on it, we have a semigroup.

Definition 1.1 (Semigroup)

A **semigroup** $(S, \cdot)$ is a set S with an associative binary operation

$$\cdot : S \times S \to S \tag{4}$$

Example 1.1 (Continuous Time Markov Chain)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(S, \mathcal{S})$ a measurable space. Then, a homogeneous continuous-time Markov chain is a stochastic process $\{X_t\}_{t\geq 0}$ taking values in S (i.e. $X_t : \Omega \to S$) satisfying the **Markov property**: for every bounded measurable f and and $t, s \geq 0$,

$$\mathbb{E}[f(X_{t+s}) \mid \{X_r\}_{r\leq t}] = \mathbb{E}[f(X_{t+s}) \mid X_t] = (P_s f)(X_t) \tag{5}$$

The set $\{P_t\}_{t\geq 0}$ with the composition operation gives us the *Markov semigroup*.

To be honest the above example is the only time I have every seen a semigroup come up, so we proceed immediately to the next structure.

Definition 1.2 (Monoid)

A **monoid** $(M, \cdot, e)$ is a semigroup with an identity element $e \in M$ such that given a $m \in M$

$$e \cdot m = m \cdot e = m \tag{6}$$

We first should ask whether the identity is unique in a monoid. It turns out it is.

Lemma 1.1 (Uniqueness of Monoid Identity)

The identity e of a monoid M is unique.

Proof. Assume not, i.e. there are 2 identities $e \neq e'$. But then

$$e = ee' = e' \implies e = e' \tag{7}$$

where the implication follows from transitivity of equivalence relations.

From set theory, we have directly worked with two examples of monoids.

Example 1.2 (Set Operations)

Let S be any nonempty set. Then $(2^S, \cup, \emptyset)$ and $(2^S, \cap, S)$ are monoids. So it seems that there are flavors of algebra that aren't really separable from set theory.

Definition 1.3 (Submonoid)

Given a monoid $(M, *)$, let $M' \subset M$. If the restriction of $*$ to $M' \times M'$ is closed in M', then we can define the **submonoid** $(M', *)$.

It may seem like the identity of a submonoid must be the identity of the monoid, but this is not always the case. We may take a subset $M' \subset M$ such that $\cdot$ is closed in M' and there may be some $e' \in M', e' \neq e$ such that it acts like an identity on M'.

Example 1.3 (Identities of Submonoids May Not be the Same)

Let $(M, \times, I)$ be the monoid of 2×2 matrices over $\mathbb{R}$ with the identity matrix I, and let M' be the set of matrices of form

$$\begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} \text{ for } a \in \mathbb{R}, \qquad I' = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{8}$$

Then $(M', \times, I')$ is a submonoid with a different identity element.

Example 1.4 ($\mathbb{N}$ is a Monoid)

The natural numbers, defined here are a monoid. More specifically,

1. $(\mathbb{N}, +, 0)$ is a monoid under addition.
2. $(\mathbb{N}, \times, 1)$ is a monoid under multiplication.
3. $(2\mathbb{N}, +, 0)$ is a monoid under addition, where $2\mathbb{N}$ is the set of all even numbers.
4. $2\mathbb{N}$ cannot be a monoid since $1 \notin 2\mathbb{N}$.

Definition 1.4 (Monoid of Transformations)

Given a set S, consider the set of all functions $S^S \coloneqq \{f : S \to S\}$. Then, with function composition $\circ$, $(S^S, \circ)$ is a monoid with the identity function $e : x \mapsto x$ as the identity element. This is called the **monoid of transformations** of S.

Theorem 1.2 (Cardinality of Monoid of Transformations)

If $|S| = n$, then the monoid of transformations has cardinality n^n.

1.2 Groups

Now we look at a specific case of monoids where invertibility is defined. The existence of inverses produces a whole suite of interesting properties, as we will see.

Definition 1.5 (Group)

A **group** $(G, \cdot)$ is a set with binary operation $x \cdot y$—also written as xy—having the following properties.

1. *Closure.* $a, b \in G \implies ab \in G$[a]
2. *Associativity.* $\forall a, b, c \in G, a(bc) = (ab)c$
3. *Identity.* $\exists e \in G$ s.t. $\forall a \in G, ae = ea = a$
4. *Inverses.* $\forall a \in G\ \exists a^{-1} \in G$ s.t. $aa^{-1} = a^{-1}a = e$

The **order** of a group is the cardinality $|G|$. An **abelian group** $(G, +)$ is a group where $+$ is commutative.[b]

[a] but not necessarily $ab = ba$

[b] Note that I switched the notation from $*$ to $+$. By convention and to avoid confusion, $+$ denotes commutative operations.

This is an extremely simple structure, and the first thing we should prove is the uniqueness of the identity and inverses.

Lemma 1.3 (Uniqueness of Identity and Inverse)

The identity and the inverse is unique, and for any a, b, the equation $xa = b$ has the unique solution $x = ba^{-1}$.

Proof. Assume that there are two identities of group $(G, *)$, denoted e_1, e_2, where $e_1 \neq e_2$. According to the properties of identities, $e_1 = e_1 e_2 = e_2 \implies e_1 = e_2$.
As for uniqueness of a inverses, let a be an element of G, with its inverses a_1^{-1}, a_2^{-1}. Then,

$$aa_1^{-1} = e \implies a_2^{-1}(aa_1^{-1}) = a_2^{-1}e \tag{9}$$
$$\implies (a_2^{-1}a)a_1^{-1} = a_2^{-1} \tag{10}$$
$$\implies ea_1^{-1} = a_2^{-1} \tag{11}$$

Since the inverse is unique, we can operate on each side of the equation $xa = b$ to get $xaa^{-1} = ba^1 \implies xe = x = ba^{-1}$. Clearly, the derivation of this solution is unique since the elements that we have operated on are unique.

At this point, we can see that for each group there is a corresponding "multiplication table" defined by the operation. For example, we can create a set of 6 elements $\{r_0, r_1, r_2, s_0, s_1, s_2\}$ and define the operation $\times$ as the following.

$\times$	r_0	r_1	r_2	s_0	s_1	s_2
r_0	r_0	r_1	r_2	s_0	s_1	s_2
r_1	r_1	r_2	r_0	s_1	s_2	s_0
r_2	r_2	r_0	r_1	s_2	s_0	s_1
s_0	s_0	s_2	s_1	r_0	r_2	r_1
s_1	s_1	s_0	s_2	r_1	r_0	r_2
s_2	s_2	s_1	s_0	r_2	r_1	r_0

Figure 1: Multiplication table for some group. Note that we can only write such a table explicitly for a group of finite elements. But even for arbitrary groups, we should think of the operation completely defining a possibly "infinite" table.

It is clear that in an abelian group, the multiplication table must be symmetric across the diagonal.

Example 1.5 (Familiar Groups)

So what are some examples of groups?

1. $(\mathbb{N}, +)$ is not a group since $3 \in \mathbb{N}$ but $-3 \notin \mathbb{N}$. It is a commutative monoid.
2. $(\mathbb{N}, \times)$ is not a group but is a commutative monoid.
3. $(\mathbb{Z}, +)$ is an abelian group.
4. $(\mathbb{Z}, \times)$ is not a group.
5. $(\mathbb{Q}, +)$ and $(\mathbb{Q} \setminus \{0\}, \times)$ are both abelian groups.
6. $(\mathbb{R}, +)$ and $(\mathbb{R} \setminus \{0\}, \times)$ are both abelian groups.
7. The set of all invertible $n \times n$ matrices with matrix multiplication, denoted $(\text{GL}(\mathbb{R}^n), \times)$ is a non-abelian group.
8. The set of all functions on a given interval $[a, b]$ is abelian with respect to addition, defined as $(f + g)(x) \equiv f(x) + g(x)$.

Example 1.6 (Group of Invertible Elements of a Monoid)

Given $x \in (M, \cdot, e)$, let x be **invertible** if there exists $x^{-1} \in M$ s.t. $xx^{-1} = x^{-1}x = e$. Then, the submonoid M' of invertible elements of M is a group. This must be proved.

1. *Closure.* If $x, y \in M'$, then $x^{-1}, y^{-1} \in M'$ since $(x^{-1})^{-1} = x$. Therefore $y^{-1}x^{-1} = (xy)^{-1} \in M'$, and so $xy \in M'$.
2. *Identity.* $e^{-1} = e$ so $e \in M'$.
3. *Inverses.* Exists by definition.
4. *Associativity.* Is inherited from associativity of $\cdot$ in M.

Let's prove a little more about groups so that we have more tools for manipulation.

Lemma 1.4 (Properties of Group Operation)

Given $a, b, c \in G$,

1. $ab = cb \implies a = c$.
2. $\forall a \in G$, $(a^{-1})^{-1} = a$.
3. $(ab)^{-1} = b^{-1}a^{-1}$.

Proof. TBD.

Theorem 1.5

Given group G, $(ab)^2 = a^2b^2$ for all $a, b \in G$ iff G is abelian.

Definition 1.6 (Subgroup)

Given group $(G, *)$, a **subgroup** $(H, *)$ is a group such that $H \subset G$. H is called a **proper subgroup** if $H \subsetneq G$.

Theorem 1.6

If $H, K \subset G$ are subgroups, then $H \cap K$ is a subgroup.

Finally we end with an analogous result of the monoid of transformations. The problem with these transformations is that they may not be invertible, but if they are, i.e. bijective, then we can endow them with a

group structure.

Definition 1.7 (Group of Transformations)

Given a set S, $\text{Sym}(S)$ is the group of bijective maps $f : S \to S$ with composition as the operator. This is also called the **symmetric group** of S.

Lemma 1.7 (Cardinality of Group of Transformations)

If S has cardinality n, then the order of $\text{Sym}(S)$ is $n!$.

1.3 Group Homomorphisms

At this point, we would like to try and classify groups (e.g. can we find *all* possible groups of a finite set?). But consider the two groups.

+	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

+	a	b	c
a	a	b	c
b	b	c	a
c	c	a	b

Figure 2: Two isomorphic groups.

These groups have different elements, but the operation behaves in exactly the same way between them (it may be a little harder if I relabeled the elements or permuted the rows/columns). Since we can trivially make arbitrary sets there really isn't much meaning to having two versions of the same group (at least in the algebraic sense). Therefore, these groups should be labeled "equivalent" in some way, and we will precisely define this notion now.

Definition 1.8 (Group Homomorphism)

Let $(G, \circ)$ and $(H, *)$ be two groups. The mapping $f : (G, \circ) \longrightarrow (H, *)$ is a **group homomorphism** if for all $a, b \in G$,

$$f(a \circ b) = f(a) * f(b) \tag{12}$$

Furthermore,

1. A **group isomorphism** is a bijective group homomorphism, and we call groups M, N **isomorphic**, denoted $M \simeq N$, if there exists an isomorphism between them.
2. An **endomorphism** is a homomorphism from a group to itself.
3. An **automorphism** is a isomorphism from a group to itself.

It turns out that from the simple property that $f(ab) = f(a)f(b)$, it also maps identities to identities, and inverses to inverses!

Lemma 1.8 (Homomorphisms Maps Identities/Inverses to Identities/Inverses)

Given a homomorphism $f : (G, *) \to (H, \times)$ and $a \in G$,

$$f(e_G) = e_H, \qquad f(a^{-1}) = f(a)^{-1} \tag{13}$$

Proof. Let $a \in G$. Then

$$f(a) = f(ae_G) = f(a)f(e_G) \implies e_H = f(a)^{-1}f(a) = f(a)^{-1}f(a)f(e_G) = f(e_G) \tag{14}$$

To prove inverses, we see that

$$f(a)f(a^{-1}) = f(aa^{-1}) = f(e_G) = e_H \tag{15}$$

from above, and this implies that $f(a^{-1}) = f(a)^{-1}$. We can also do this with right hand side multiplication.

Example 1.7 (Exponential Map)

The map $a \mapsto 2^a$ is an isomorphism between $(\mathbb{R}, +)$ and $(\mathbb{R}^+, \times)$ since

$$2^{a+b} = 2^a \times 2^b \tag{16}$$

which is proved in my real analysis notes when constructing the exponential map on the reals.

Example 1.8 (Determinant)

The determinant $\det : \mathrm{GL}_n(\mathbb{F}) \to \mathbb{F}^*$ is a homomorphism because of the product rule for determinants.

Example 1.9 (Projection onto Unit Circle)

Given $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$ with $\times$ and $S^1 = \{x \in \mathbb{C} \mid |x| = 1\}$ (which is a group under multiplication), the map $f : \mathbb{C}^* \to S^1$ defined $f(z) = z/|z|$ is a group homomorphism since

$$f(z_1 z_2) = \frac{z_1 z_2}{|z_1 z_2|} = \frac{z_1 z_2}{|z_1||z_2|} = f(z_1)f(z_2) \tag{17}$$

Definition 1.9 (Kernel)

Given group homomorphism $f : G \to H$, the **kernel** of f is the preimage of the identity.

$$\ker(f) \coloneqq \{g \in G \mid f(g) = e_H\} \tag{18}$$

Theorem 1.9 (Images and Kernels of Group Homomorphisms)

If $f : G \to H$ is a group homomorphism, then
1. $\ker(f)$ is a subgroup of G.
2. $\mathrm{Im}(f)$ is a subgroup of H.
3. f is injective iff $\ker(f) = \{e_G\}$.

Proof. Listed.
1. To show closed, consider $a, b \in \ker(f)$. Then $f(ab) = f(a)f(b) = e_H e_H = e_H \implies ab \in \ker(f)$. Since $f(e_G) = e_H$, $e_G \in \ker(f)$. If $a \in \ker(f)$, then $f(a^{-1}) = f(a)^{-1} = e_H^{-1} = e_H \implies a^{-1} \in \ker(f)$. Finally associativity follows from associativity of the supgroup.
2. TBD
3. We prove bidirectionally.

(a) ($\to$). Since f is injective, $f(a) = f(b) \implies a = b$. Let $a \in \ker(f)$. Then $f(a) = e_H$, and so $f(e_G) = e_H = f(a)$. By injectivity, $a = e_G$, and so $\ker(f) = \{e_G\}$.
(b) ($\leftarrow$). Let $a, b \in G$ s.t. $f(a) = f(b)$. Then $f(a)f(b)^{-1} = e_H \implies af(a)f(b^{-1}) = f(ab^{-1}) = e_H \implies ab^{-1} \in \ker(f)$. But by hypothesis $\ker(f) = \{e_G\} \implies ab^{-1} = e_G \implies a = b$.

Theorem 1.10 (Compositions of Group Homomorphisms)

Compositions of group homomorphisms are group homomorphisms.

Now let's focus more on isomorphisms, which we can interpret as a "renaming" of the elements. Not only does it rename the elements, but it preserves all the algebraic properties of the group and each element.

Theorem 1.11 (Properties of Group Isomorphisms)

If $f : G \to H$ is a group isomorphism, then
1. f^{-1} is also a group isomorphism.
2. G is abelian $\implies$ H is abelian.

Proof. Listed.
1. Since f is bijective by definition, f^{-1} is well-defined and bijective as well. Now we show that f^{-1} is a group homomorphism. Given $c, d \in H$, take
$$f\big(f^{-1}(c), f^{-1}(d)\big) = f\big(f^{-1}(c)\big)\, f\big(f^{-1}(d)\big) = cd \tag{19}$$
where the first equality follows since f is a homomorphism, and the second since f^{-1} is the inverse mapping. Now mapping both sides through f^{-1}, we get
$$f^{-1}(c)f^{-1}(d) = f^{-1}(cd) \tag{20}$$
and so f^{-1} is a homomorphism.
2. Let $c, d \in H$. Then $c = f(a), d = f(b)$ for some $a, b \in G$, and so $cd = f(a)f(b) = f(ba) = f(b)f(a) = dc$.

A trivial example is the identity map, which is an automorphism. But can we generalize this a bit better?

Theorem 1.12 (Conjugate Shift is an Automorphism)

Let G be a group with $a \in G$. Then the following, called a **conjugate shift**, is an automorphism on G.
$$\phi : G \longrightarrow G,\ \phi(x) = axa^{-1} \tag{21}$$

Proof. The map $\psi : G \longrightarrow G,\ \psi(x) = a^{-1}xa$ is clearly the inverse of ϕ, with $\phi\psi = \psi\phi = I$ for all $x \in G \implies \phi$ is bijective. Secondly, $\phi(x)\phi(y) = axa^{-1}aya^{-1} = a(xy)a^{-1} = \phi(xy) \implies \phi$ preserves the group structure.

1.4 Group Presentations

A group G may be very abstract and complicated, and so working with all its elements can be a bit painful. It would be more useful to work with a smaller subset S of G that can completely characterize G.[1] We would like to formalize this notion, which will be very useful later on. For now, let's start off with a simple

[1] Note that this is similar to the basis that generates a topology.

element $a \in G$, and perhaps we can consider the elements

$$\dots, a^{-2}, a^{-1}, a^0 = e, a^1, a^2, \dots \tag{22}$$

However, there are two interpretations to a^{-2} is it the inverse of a^2 or $a^{-1}a^{-1}$? It turns out that these are equivalent.

Lemma 1.13 (Power to an Integer is Well-Defined)

For all $n \in \mathbb{N}$,

$$(a^{-1})^n = (a^n)^{-1} \tag{23}$$

Proof. We prove by induction on n. It is trivially true for $n = 1$. Now given that it is true for some $n \in \mathbb{N}$, we have

$$(a^{-1})^{n+1} = (a^{-1})^n a^{-1} = (a^n)^{-1}a^{-1} = (aa^n)^{-1} = (a^{n+1})^{-1} \tag{24}$$

Therefore it makes sense to just write it as a^{-n}. It may or may not be the case that a may cycle back to itself for some n, i.e. $a = a^n$.

Definition 1.10 (Order of an Element)

The **order** of a group element $a \in G$ is the minimum number $n \in \mathbb{N}$ s.t. $a = a^n$, denoted $|a|$ or $\text{ord}(a)$.[a]

[a] Note that this is different from the order of a group. This is confusing but is the convention.

Let's pause for a bit and focus on the order. The order of an element a tells us how a "behaves" in the broader group. This means that when mapped through an isomorphism, it should behave similarly, i.e. the order shouldn't change. This gives us a nice way to check if two groups cannot be isomorphic, but the converse is not necessarily true in general!

Theorem 1.14 (Preservation of Order in a Group Homomorphism)

If f is a group isomorphism, then $\forall a \in G$ $\text{ord}(a) = \text{ord}(f(a))$.

Proof.

Now we come back to group presentations. The set of all multiples of a may or may not be the group, but if we take a certain subset of these elements and take all multiples of all combinations of them, we may have better coverage of the group.

Definition 1.11 (Word)

A **word** is any written product of group elements and inverses. They are generally in the form

$$s_1^{\epsilon_1} s_2^{\epsilon_2} s_3^{\epsilon_3} ... s_k^{\epsilon_k}, \text{ where } e_i \in \mathbb{Z} \tag{25}$$

e.g. given a set $\{x, y, z\}$, $xy, xz^{-1}yyx^{-2}, ...$ are words.

Definition 1.12 (Generating Set)

The **generating set** $\langle S \rangle$ of a group G is a subset of G such that every element of the group can be expressed as a word of finitely many elements under the group operations. The elements of the generating set are called **generators**.

Definition 1.13 (Group Presentations)

The **free group** F_S over a given set S consists of all words that can be built from elements of S. Often with this generating set S, we have a set of relations R that tell us which elements are equal. The **group presentation** writes both S and R in the form

$$\langle S \mid R \rangle \tag{26}$$

Theorem 1.15

If every element other than the identity has order 2, then G is abelian.

With these group presentations we can start identifying specific groups. Let's start with the simplest group with one generator and zero/one relation: the cyclic group.

Definition 1.14 (Cyclic Group)

A **cyclic group** is a group generated by a single element.

1. In an infinite cyclic group, there is no relation and we write
$$Z \coloneqq \langle a \rangle = \{\dots, a^{-2}, a^{-1}, e, a, a^2, \dots\} \tag{27}$$
2. In a finite cyclic group, there exists a $n \in \mathbb{N}$ such that $a^n = e$ and we write
$$Z_n \coloneqq \langle a \mid a^n = e \rangle = \{e, a, a^2, \dots, a^{n-1}\} \tag{28}$$

Example 1.10 (Cyclic Groups)

Here are some examples of cyclic groups.

1. $(\mathbb{Z}_n, +)$, the integers mod n, is a cyclic group of order n, generated by 1.[a]
2. The nth roots of unity in $\mathbb{C}$ is a cyclic group of order n, generated by the counterclockwise rotation $e^{2\pi/n}$.
3. The set of discrete angular rotations in $SO(2)$, in the form of
$$R = \left\{ \begin{pmatrix} \sin\theta & \cos\theta \\ \cos\theta & -\sin\theta \end{pmatrix} \;\middle|\; \theta \in \left\{ \frac{2\pi}{n} k \right\}_{k=0}^{n-1} \right\} \tag{29}$$
4. $(\mathbb{Z}, +)$ is an infinite cyclic group.

[a] In fact, the generator of $\mathbb{Z}_n$ can be any integer relatively prime to n and less than n.

That's really it for cyclic groups, and to make things simpler, there is a complete characterization of them.

Theorem 1.16 (Cyclic Groups are Unique up to Order)

Given a cyclic group, Z or Z_n
1. If it is finite, then $(Z_n, +) \simeq (\mathbb{Z}_n, +) \simeq \langle 1 \rangle$.
2. If it is infinite, then $(Z, +) \simeq (\mathbb{Z}, +) \simeq \langle 1 \rangle$.

Proof.

Therefore, we have completely characterized all cyclic groups! Furthermore, cyclic groups are contained in the sense that any subgroup is also a cyclic group. So you won't find any weird groups embedded in cyclic groups; you can safely assume that they are all cyclic. The proof for this is quite a useful technique, where we try to arrive at a contradiction between some minimally chosen k and the remainder r that must be less than k.

Theorem 1.17 (Subgroups of Cyclic Groups)

Any subgroup of a cyclic group is cyclic.

Proof. Let $G = \langle a \rangle$ be a cyclic group. Then given a subgroup H, we must have $e \in H$. If there are no other elements we are done, and if there are extra elements then let $k \in \mathbb{N}$ be the smallest natural (which exists due to the well-ordering principle) such that $a^k \in H$. Now we claim that $H = \langle a^k \rangle$. Given any $a^n \in H$, we can use Euclidean algorithm on the integers to write $n = qk + r$ for $0 \leq r < k$. Therefore,

$$a^n = a^{qk+r} = (a^k)^q \cdot a^r \implies a^r = a^n (a^k)^{-q} \tag{30}$$

$$\implies a^r \in H \tag{31}$$

but this contradicts the fact that k is minimal, and so $r = 0$. This means that $a^n = (a^k)^q$ and so a^n is a multiple of a^k.

Example 1.11 (Integers to Even Integers)

Let $2\mathbb{Z}$ denote the set of all even integers with addition. Then we can verify that this is a group, and

$$\mathbb{Z} \simeq 2\mathbb{Z} \tag{32}$$

Theorem 1.18 (Homomorphisms between Cyclic Groups)

There are precisely $\gcd(n, m)$ homomorphisms $f : Z_n \to Z_m$.

Proof.

The next type of group we will focus on is the dihedral group. These are usually introduced as the symmetry group (group of rotations and flips you can do on a polygon) to preserve its symmetry. However, it seems a bit disconnected with cyclic groups and group presentations, so I introduce it in the following way. Once I define it, I connect to its geometric interpretations in the following examples.

Definition 1.15 (Dihedral Group)

The **Dihedral Group** of order $2n$ is the group

$$\text{Dih}(n) \coloneqq \langle r, f \mid r^n = f^2 = e, rfr = f \rangle \tag{33}$$

To parse this definition a bit, note that the relation $r^n = e$ behaves like a cyclic group of order n, and so we can interpret these as rotations of an object by $2\pi/n$. The second is that $f^2 = e$ is also a cyclic group of order 2, but it behaves more like a flip in that if you flip twice, you get back to the original. With these relations, we can think of the Dihedral group as having two "copies" of cyclic groups that have some extra properties.

Finally, the relation $rfr = f$ is a bit harder to parse, but it just means that a rotation, then flip, then rotation (which rotates backwards since we flipped), is equal to flipping once. Symbolically, this relation allows us to "push" all of the flips to the back.

$$fr = r^n fr = r^{n-1} f \tag{34}$$

Perhaps a slightly more complicated example for $n = 5$.

$$fr^3 f^3 r = fr^3 fr = fr^2 f = r^5 fr^2 f = r^4 frf = r^3 f^2 = r^3 \tag{35}$$

and after this the relation $r^n = f^2 = e$ allows us to cancel some out.

Example 1.12 (Dihedral Group of Order 4, aka Klein-4 Group)

We use the following group presentation to write the dihedral group of order 4. However, we can relabel them to get a simpler table.

	e	r	f	rf
e	e	r	f	rf
r	r	e	rf	f
f	f	rf	e	r
rf	rf	f	r	e

$\cdot$	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

Figure 3: Cayley multiplication table for the Klein 4-group.

It can be described as the symmetry group of a non-square rectangle. With the three non-identity elements being horizontal reflection, vertical reflection, and 180-degree rotation.

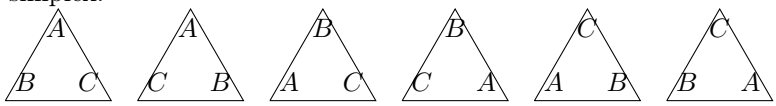
Example 1.13 (Dihedral Group of Order 6)

The group of rotations and flips you can do on a equilateral triangle is called the Dihedral Group Dih(3). It is not abelian.

	e	r	r^2	f	rf	r^2f
e	e	r	r^2	f	rf	r^2f
r	r	r^2	e	rf	r^2f	f
r^2	r^2	e	r	r^2f	f	rf
f	f	r^2f	rf	e	r^2	r
rf	rf	f	r^2f	r	e	r^2
r^2f	r^2f	rf	f	r^2	r	e

Figure 4: Multiplication table for D_3 using simplified notation.

Dih(3) is the group of all rotations and reflections that preserve the structure of the equilateral triangle in $\mathbb{R}^2$, a regular 2-simplex.



Example 1.14 (Dihedral Group of Order 8)

The group of rotations and reflections that preserve the structure of a square in $\mathbb{R}^2$. is called the Dihedral Group Dih(4).

	e	r	r^2	r^3	f	rf	r^2f	r^3f
e	e	r	r^2	r^3	f	rf	r^2f	r^3f
r	r	r^2	r^3	e	rf	r^2f	r^3f	f
r^2	r^2	r^3	e	r	r^2f	r^3f	f	rf
r^3	r^3	e	r	r^2	r^3f	f	rf	r^2f
f	f	r^3f	r^2f	rf	e	r^3	r^2	r
rf	rf	f	r^3f	r^2f	r	e	r^3	r^2
r^2f	r^2f	rf	f	r^3f	r^2	r	e	r^3
r^3f	r^3f	r^2f	rf	f	r^3	r^2	r	e

Figure 5: Multiplication table for D_4 using simplified notation.

Note that this is **not** the same as the symmetry group of the regular tetrahedron!

Following this pattern, we can extrapolate to find that the Dihedral group is a symmetry group.

Theorem 1.19 (Dihedral Groups as Symmetry Groups)

Dih(n) is similarly the group of all rotations and reflections that preserve the structure of a regular n-gon in $\mathbb{R}^2$.

Example 1.15 (Groups of Order 3)

Dih(3) $\simeq S_3$, since permutations of the vertices of a triangle are isomorphic to a permutations of a 3-element set.

Theorem 1.20 (Tip)

To prove a group homomorphism, show that every element of G and H can be written as a word of certain g_i's in G and then h_i's in H, and map the g_i's to h_i's.

1.5 Symmetric and Alternating Groups

We have seen the natural construction of the symmetric group of a set as the set of bijective transformations. Now the reason that symmetric groups are nice is that we can embed a group into its symmetric group.

Theorem 1.21 (Cayley's Theorem)

This applies for both monoids and groups.

1. Any monoid is isomorphic to a monoid of transformations, i.e. there exists an injective monoid homomorphism
$$f : M \to M^M \tag{36}$$
2. Any group is isomorphic to a group of transformations, i.e. there exists an injective group homomorphism
$$f : G \to \text{Sym}(G) \tag{37}$$

Proof. Let $(M, \cdot, 1)$ be a monoid. Then we will construct a homomorphism $f : M \to M^M$, the monoid of transformations from M to itself. For any $a \in M$, we define the *left translation* $a_L : x \mapsto ax$. We claim that the set $M' \coloneqq \{a_L \in M^M \mid a \in M\}$ is indeed a monoid.

1. *Closure.* Given $a, b \in M$, $ab \in M$ and so $ab_L \in M'$. But $(ab_L)(x) = (ab)x = a(bx) = a_L(bx) = a_L(b_L(x)) = (a_L \circ b_L)(x)$, so $ab_L = a_L \circ b_L$.
2. *Identity.* $e \in M \implies e_L \in M$ where $e_L : x \mapsto x$.

Next we claim that is is an isomorphism.

1. This is a homomorphism due to the closure and identity properties proved above.
2. It is injective since given $a \neq b$ in M, a_L and b_L acts on the identity in different ways $a_L(e) = a \neq b = b_L(e)$, so $a_L \neq b_L$.
3. It is surjective by definition.

We have proved for monoids. For groups, we have the additional assumption that inverses exist in G, and we must prove that the set of left translations G' is indeed a group. It suffices to prove that inverses exist in G'. Given $a \in G$, $a_L \in G'$. But $a^{-1} \in G$ since G is a group, and so $a_L^{-1} \in G'$ as well. We can see that

$$(a_L^{-1} a_L)(x) = (a^{-1}a)x = ex = x \tag{38}$$
$$(a_L a_L^{-1})(x) = (aa^{-1})x = ex = x \tag{39}$$

and so indeed $(a^{-1})_L = (a_L)^{-1}$. From this additional fact all the rest follows exactly as for monoids.

Corollary 1.22 (Cayley)

Every group G is isomorphic to a subgroup of its symmetric group.

Now we limit our scope to only finite sets, i.e. finite symmetric groups, which are often called **permutation** groups. For such finite sets the labeling does not matter since such groups are always isomorphic, so we can say $S = \{1, 2, \dots, n\}$.

Theorem 1.23 (Symmetric Group as a Symmetry Group)

The symmetric group S_n is isomorphic to the symmetry group of the n-simplex in $\mathbb{R}^{n-1}$.

Proof.

Now armed with group presentations and generating sets, let attempt to find a group presentation for a permutation group. Given set $S = \{1, 2, \dots, n\}$, a permutation $\gamma \in \text{Sym}(S)$ is denoted

$$\gamma = \begin{pmatrix} 1 & 2 & 3 & 4 & \dots & n-2 & n-1 & n \\ i_1 & i_2 & i_3 & i_4 & \dots & i_{n-2} & i_{n-1} & i_n \end{pmatrix} \in \text{Sym}(S) \tag{40}$$

We begin by introducing a specific instance of a permutation.

Definition 1.16 (Cyclic Permutation)

A permutation is said to be **cyclic** if there exists some subset $A \subset S$ such that γ acts as

$$a_1 \mapsto a_2 \mapsto a_3 \ldots \mapsto a_k \mapsto a_1 \tag{41}$$

and leaves the rest unchanged. The notation for this is

$$\begin{pmatrix} a_1 & a_2 & \ldots & a_k \end{pmatrix} \in \text{Sym}(S) \tag{42}$$

A cycle acting on a subset of 2 elements, i.e. a swap of two elements, is called a **transposition**. Two cyclic rotations γ_1, γ_2 are **disjoint** if the subsets that they act on are disjoint: $A \cap B = \emptyset$.

Example 1.16 (Some Cyclic Permutations)

This notation can be a bit weird, so let's give some simple examples.

1. (12) is a mapping $1 \to 2,\ 2 \to 1$.
2. (123) is a mapping $1 \to 2,\ 2 \to 3,\ 3 \to 1$.
3. $(123)(45)$ is a mapping $1 \to 2,\ 2 \to 3,\ 3 \to 1,\ 4 \to 5,\ 5 \to 4$.

The reason that cyclic permutations are so important is that they are the building blocks of regular permutations.

Theorem 1.24 (Cycle Decomposition of Permutations)

Every element in S_n except the identity element can be written uniquely (up to order) as the product of disjoint cycles.

Proof. We can compute $\gamma(1), \gamma^2(1), \ldots$. Since $S = \{1, \ldots, n\}$ is finite, there is some smallest positive natural k s.t.
$gamma^k(1) = 1$. This yields a k-cycle. Now remove the numbers $1, \pi(1), \ldots, \gamma^{k-1}(1)$ and continue the process. Since S is finite this must terminate, and we have such a decomposition. Proof of uniqueness omitted for now, but this whole theorem can be proved using proof by strong induction.

Example 1.17 (Cyclic Decompositions)

For the following permutation

$$\gamma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 6 & 5 & 4 & 8 & 2 & 7 & 1 \end{pmatrix} = (7)(4)(26)(1358) \tag{43}$$

One of the rewards of cycle decompositions is that we can easily compute the effect of conjugation in S_n.

Lemma 1.25 (Conjugation is Easy with Cycle Notation)

Given a k-cycle $\gamma = (i_1, i_2, \ldots, i_k) \in S_n$ and any permutation $\sigma \in S_n$, we have

$$\sigma\pi\sigma^{-1} = \begin{pmatrix} \sigma(i_1) & \sigma(i_2) & \ldots & \sigma(i_k) \end{pmatrix} \tag{44}$$

This shows that the cyclic permutations actually form a generating set of S_n. But can we do better? The answer is yes: there is a more minimal generating set.

Corollary 1.26 (Transposition Decomposition of Permutations)

The set of all transpositions forms a generating set of S_n.

Proof. It suffices to prove that the cycles can be decomposed into transpositions. Indeed, we can just write out by hand

$$\begin{pmatrix} 1 & 2 & \dots & k \end{pmatrix} = \begin{pmatrix} 1 & k \end{pmatrix} \begin{pmatrix} 1 & k-1 \end{pmatrix} \dots \begin{pmatrix} 1 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 \end{pmatrix} \tag{45}$$

which by relabeling generalizes for those of form $(i_1 \dots i_k)$.

Recognizing that the set of transpositions is the generating set of the permutation group, we must prove a few more statements before constructing the alternating group. One such fact is that transpositions allow us talk about the parity of an arbitrary permutation, through its signature.

Lemma 1.27 (Parity of Transpositions)

Every permutation can be written as the product of either an even number or an odd number of transpositions, but not both.

Proof.

Now that this is established, the following is well-defined.

Definition 1.17 (Signature)

The **signature** of a permutation is a homomorphism

$$\text{sgn} : S_n \longrightarrow \{1, -1\} \tag{46}$$

Lemma 1.28

The signature of a permutation changes for every transposition that is applied to it.

Now we are ready to introduce another fundamental type of group.

Definition 1.18 (Alternating Group)

The **alternating group** is the kernel of the signature homomorphism $\text{sgn} : S_n \to \{\pm 1\}$.

$$A_n \coloneqq \ker \text{sgn} \tag{47}$$

It is the set of even permutations with order $n!/2$.[a]

[a] Note that the set of odd permutations do not form a group, since the composition of two odd permutations (each having signature -1 is an even permutation.

This construction might seem arbitrarily specific to study so early into algebra, but as we will see later, we will find that that for $n \geq 5$, they will be simple groups that can't be decomposed and therefore fundamental in a sense.

In lecture, we talked about the number of all finite set is e. Since $n!$ is the order of permutation groups, i.e. the order of automorphism groups, we can sum their inverses over all $n \in \mathbb{N}$ to get e.

1.6 Group Actions

We have studied the general properties of groups, but historically group theory arose from the study of transformation groups (which is why I also introduced is so early on). These transformation groups can be thought of as an abstract group it itself, but another way to interpret it is to see how it *acts* on a set.

Definition 1.19 (Group Action)

Let G be a group, S a set. Then, a (left) group action of G on S is a function

$$\sigma : G \times S \to S, \qquad \sigma(g, a) = g \cdot a \tag{48}$$

satisfying two axioms.

1. *Identity.* $\forall a \in S, \sigma(e, a) = a$.
2. *Compatibility.* $\forall g, h \in G$ and $\forall x \in X, \sigma(gh, x) = \sigma(g, \sigma(h, x))$.

The group G is said to **act on** S, and the evaluation $\sigma(g, a)$ can be interpreted as the result after transforming a through g.

Theorem 1.29 (Group Action as a Homomorphism onto the Symmetric Group)

We have the immediate facts.

1. For a fixed $g \in G$, the group action $\sigma_g(s) \coloneqq \sigma(g, s) : S \to S$ is a bijection, i.e. an element of $\text{Sym}(S)$. The inverse is the function mapping $x \mapsto \sigma(g^{-1}, x)$.
2. The map from G to $\text{Sym}(S)$ defined by $g \mapsto \sigma_g$ is a homomorphism.

Proof.

Example 1.18 (Permutations and Dihedral Groups as Group Actions)

In fact, we have seen two concrete examples of such group actions.

1. The permutation group acts on the set $S = \{1, \dots, n\}$ by permuting its elements. It also acts on a set of n-simplexes by rotating/flipping them.
2. The dihedral group acts on the set of regular n-gons by rotating/flipping them.

1.7 Exercises

2 Subgroups

We have seen a few examples of subgroups, but we will heavily elaborate on here. We know that given a set, we can define an equivalence relation on it to get a quotient set. Now if we have a group, defining any such equivalence relation may not be compatible with the group structure. Therefore, it would be nice to have some principles in which we can construct such compatible equivalence classes, i.e. through a **congruence relation** that preserves the operations.

We introduce some standard notation.

Definition 2.1 (Subgroup of Integer Multiples)

The set $k\mathbb{Z}$ is the set of all integer multiples of k. This is a group under addition.

2.1 Cosets

Fortunately, we can do such a thing by taking a subgroup $H \subset G$ and "shifting" it to form the cosets of G, which are the equivalence classes.

Definition 2.2 (Coset)

Given a group G, $a \in G$, and subgroup H,
1. A **left coset** is $aH \coloneqq \{ah \mid h \in H\}$.
2. A **right coset** is $Ha \coloneqq \{ha \mid h \in H\}$.
3. When G is abelian, the **coset** is denoted $a + H$.

With this, we can take arbitrary elements $a, b \in G$ and determine if they are in the same coset as such. Since $a \in aH$, $b \in aH$ iff $b = ah$ for some $h \in H$. Therefore, we have the equivalence relation.

$$a \equiv b \pmod{H} \iff a = bh \text{ for some } h \in H \tag{49}$$

Proof. We show that this indeed forms an equivalence class.
1. *Reflexive.* $a \equiv a \pmod{H}$ since $e \in H \implies a = ae$.
2. *Symmetric.* Let $a \equiv b \pmod{H}$. Then $a = bh$ for some $h \in H$, but since H is a group, $h^{-1} \in H \implies ah^{-1} = b \implies b \equiv a \pmod{H}$.
3. *Transitive.* Let $a \equiv b \pmod{H}$ and $b \equiv c \pmod{H}$. Then $a = bh$ and $b = ch'$ for some $h, h' \in H$. But then
$$a = bh = (ch')h = c(h'h) \tag{50}$$
where $h'h \in H$ due to closure.

Note that a coset is *not* a subgroup. It is only the case that $eH = H$ is a subgroup, but for $a \neq e$, aH does not even contain the identity. We should think of a coset as a *translation* of the subgroup H.

Example 2.1 (Familiar Cosets)

Here are some examples. Note that all it takes is to find *some* subgroup, and the cosets will naturally pop up.
1. Let $H = 2\mathbb{Z} \subset (\mathbb{Z}, +)$ be the even integers. Then $0 + H$ and $1 + H$ are the even and odd integers, respectively.
2. Let $H = \{e, f\} \subset \text{Dih}(3)$. Then
$$H = \{e, f\}, rH = \{r, rf\}, r^2H = \{r^2, r^2f\} \tag{51}$$
are the cosets.

With this partitioning scheme in mind, the following theorem on the order of such groups becomes very intuitive, and has a lot of consequences.

Theorem 2.1 (Lagrange's Theorem)

Let G be a finite group and H its subgroup. Then

$$|G| = [G : H]|H| \tag{52}$$

where $[G : H]$, called the **index of** H, is the number of cosets in G. Therefore, the order of a subgroup of a finite group divides the order of the group.

Proof. The union of the $[G : H]$ disjoint cosets is all of G. On the other hand, every H is in one-to-one correspondence with each coset aH, so every coset has $|H|$ elements. Therefore, there are $[G : H]|H|$ elements altogether.

Therefore, Lagrange's theorem says that *given* that you find a subgroup, the order of the subgroup must divide the order of G. However, that doesn't mean that such a subgroup may even exist. For example, there is a group of order 12 having no subgroup of order 6.

Corollary 2.2

The order of any element of a finite group divides the order of the group.

Proof. Take any $a \in G$ and construct the cyclic subgroup $\langle a \rangle \subset G$. Then by Lagrange's theorem, $|a| = |\langle a \rangle|$ divides $|G|$.

Corollary 2.3

Every finite group of a prime order is cyclic.

Proof. Let $a \in G$ be any element other than the identity e, and consider $\langle a \rangle \subset G$. The order must divide $|G|$ which is prime, so $|a| = 1$ or $|G|$. But $|a| \neq 1$ since we did not choose the identity, so $|a| = |G| \implies \langle a \rangle = G$.

Corollary 2.4

If $|G| = n$, then for every $a \in G$ $a^n = e$.

Proof. Let $|a| = k$. Then $k \mid n$, and so $a^n = a^{kl} = (a^k)^l = e^l = e$.

Corollary 2.5 (Fermant's Little Theorem)

Let p be a prime number. The multiplicative group $\mathbb{Z}_p \setminus \{0\}$ of the field $\mathbb{Z}_p$ is an abelian group of order $p - 1 \implies g^{p-1} = 1$ for all $g \in \mathbb{Z}_p \setminus \{0\}$. So,

$$a^{p-1} \equiv 1 \iff a^p \equiv a \pmod{p} \tag{53}$$

We can generalize this.

Definition 2.3 (Euler's Totient Function)

Euler's Totient Function, denoted $\varphi(n)$, consists of all the numbers less than or equal to n that are coprime to n.

Theorem 2.6 (Euler's Theorem)

For any n, the order of the group $\mathbb{Z}_n \setminus \{0\}$ of invertible elements of the ring $\mathbb{Z}_n$ equals $\varphi(n)$, where φ is Euler's totient function. In other words with $G = \mathbb{Z}_n \setminus \{0\}$,

$$a^{\varphi(n)} \equiv 1 \pmod{n}, \quad \text{where } a \text{ is coprime to } n \tag{54}$$

Example 2.2

In $\mathbb{Z}_{125} \setminus \{0\}$, $\varphi(125) = 125 - 25 = 100 \implies 2^{100} \equiv 1 \pmod{125}$

2.2 Normal Subgroups

By introducing cosets, we have successfully constructed an equivalence relation on G. This set of cosets is indeed a partition of G, but we would like to endow it with a group structure that respects that of G. That is, let $a, b \in G$ and its corresponding cosets be aH, bH. Then, we would like to define an operation $\cdot$ on the cosets such that

$$(aH) \cdot (bH) \coloneqq (ab)H \tag{55}$$

That is, we would like to upgrade the equivalence relation to a *congruence relation*. If we try to show that this is indeed a well-defined operation, we run into some trouble. Suppose $aH = a'H$ and $bH = b'H$. Then with our definition, we should be able to derive that $(aH)(bH) = (a'H)(b'H)$ through the equation

$$(aH)(bH) = (ab)H = (a'b')H = (a'H)(b'H) \tag{56}$$

We have $a' = ah_1$, $b' = bh_2$, and $a'b' = abh$. Then,

$$(ab)H = (a'b')H \implies a'b' = abh \text{ for some } h \in H \tag{57}$$

$$\implies ah_1bh_2 = abh \text{ for some } h_1, h_2, h \in H \tag{58}$$

But the final statement is not true in general. In an abelian group, we could just swap h_1 and b to derive it completely, but perhaps there is a weaker condition on just the subgroup H that allows us to "swap" the two.

Definition 2.4 (Normal Subgroups)

A subgroup $N \subset G$ is a **normal subgroup** iff the left cosets equal the right cosets. That is, $\forall g \in G, h \in H$.

$$g^{-1}hg \in H \tag{59}$$

We call $g^{-1}hg$ the **conjugate** of h by g.

Example 2.3 (Normal Subgroups)

For intuition, we provide some examples of normal subgroups.

1. If G is abelian, every subgroup is normal. So $(2\mathbb{Z}, +)$ is normal, and $(\mathbb{Q}, \times) \subset (\mathbb{R}, \times)$ is also normal.
2. Given $G = (\mathbb{R} \setminus \{0\}, \times)$, let $H = (\mathbb{R}^+, \times) \subset G$ be a subgroup. Then H is normal since for any

$g \in \mathbb{R}$, g, g^{-1} are either both positive or both negative, and so $ghg^{-1} > 0 \implies ghg^{-1} \in H$. H and $(-1)H$ are two cosets of $\mathbb{R}$.

3. $\mathrm{SL}_n(\mathbb{F}) \subset \mathrm{GL}_n(\mathbb{F})$ is a normal subgroup since the determinant of the inverse is ihe inverse of the determinant, and so for any $g \in \mathrm{GL}_n(\mathbb{F})$,

$$\det(ghg^{-1}) = \det(g)\det(h) = \det(g^{-1}) = \det(g) \cdot 1 \cdot \frac{1}{\det(g)} = 1 \implies ghg^{-1} \in \mathrm{SL}_n(\mathbb{F}) \tag{60}$$

4. The subgroup $H = \{e, r^2\} \subset \mathrm{Dih}(4)$ is a normal subgroup. It is clearly a subgroup isomorphic to Z_2, and to see normality, note that r^2 commutes with any $g = r^n \in \mathrm{Dih}(4)$. If g contains a flip, then we can just check the 4 cases knowing that $fr = r^3 f$.

$$fr^2 f^{-1} = fr^2 f = (fr)(rf) = r^3 frf = r^3 r^3 f^2 = r^2 \tag{61}$$

$$(rf)r^3(rf)^{-1} = \ldots = r^2 \tag{62}$$

Therefore $\mathrm{Dih}(4)/H$ has order 4, which means it must be isomorphic to either the cyclic group or the Klein 4 group. It turns out it's the Klein 4 group.

5. The subgroup $H = \{e, r, r^2, r^3\} \subset \mathrm{Dih}(4)$ is a normal subgroup because

$$\underbrace{(f^j r^i)}_{g}\underbrace{(r^l)}_{h}\underbrace{(r^{-i} f^{-j})}_{g^{-1}} = f^j r i + l - i f^{-j} \tag{63}$$

$$= f^j r^l f^{-j} \tag{64}$$

$$= f^j r^l f_j \tag{65}$$

$$= r^{l+3j} \tag{66}$$

where we used the fact that $frf = r^3 = r^{-1}$ in the penultimate step. So $|\mathrm{Dih}(4)/H| = 2 \implies \mathrm{Dih}(4)/H \simeq Z_2$ with generator fh.

Example 2.4 (Subgroups that are Not Normal)

Here are some subgroups that are not normal.

1. Given $G = \mathrm{Dih}(3)$, $H = \{e, f\}$ is not normal since $rfr^{-1} = rfr^2 = r^2 f \notin H$.
2. The subgroup

$$H = \left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \;\middle|\; ac \neq 0 \right\} \subset \mathrm{GL}_2(\mathbb{R}) \tag{67}$$

is not normal since

$$h = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in H, a = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \in \mathrm{GL}_2(\mathbb{R}) \implies aha^{-1} = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix} \notin H \tag{68}$$

Finally, we present some relevent results of alternating subgroups.

Theorem 2.7 (Alternating Group is Normal in Symmetric)

A_n is a normal subgroup of S_n, of index[a] 2.

[a] i.e. the number of cosets

Proof.

Lemma 2.8 (Cycles in Alternating Group)

We have the following.
1. Every element of A_n can be written as the product of 3-cycles.
2. If $n \geq 4$, H is a normal subgroup of A_n, and H contains one 3-cycle, then $H = A_n$.

Proof. Since we've proved that every permutation is the product of transpositions, it suffices to prove that the product of two transpositions can be written as the product of 3-cycles. We check this case by case, where distinct symbols represent distinct values.
1. $(\alpha\ \beta)(\gamma\ \delta) = (\alpha\ \beta\ \gamma)(\beta\ \gamma\ \delta)$
2. $(\alpha\ \beta)(\alpha\ \gamma) = (\alpha\ \gamma\ \beta)$
3. $(\alpha\ \beta)(\alpha\ \beta) = e$

Therefore every even permutation is the product of 3-cycles.

Definition 2.5 (Simple Group)

A **simple group** is a group with no proper normal subgroup. That is, the only normal subgroups are the trivial group and itself.

Theorem 2.9 (Alternating Groups are Simple)

For $n \geq 5$, A_n is a simple group.

Proof. Let $H \subset A_n$ be a normal subgroup containing more than the identity. If we can find a single 3-cycle in H, then it follows from 2.8 that $H = A_n$. Let $\gamma \in H$, $\gamma \neq e$, and write $\gamma = \gamma_1 \dots \gamma_m$ as a product of disjoint cycles. We have 4 cases.

1. Let $k \geq 4$ and suppose that some factor, say γ_1 is a k-cycle. WLOG let us assume that $\gamma_1 = (1 \dots k)$. Since H is normal, $(1,2,3)\gamma(1,2,3)^{-1} \in H$ and $(1,2,3)$ commutes with all the factors of γ except γ_1 (since the cycles are disjoint and so γ_i for $i \neq 1$ does not contain $1,2,3$). Thus letting
$$\sigma = (1,2,3)\gamma(1,2,3)^{-1} = (2,3,1,4,\dots,k)\gamma_2 \dots \gamma_m \in H \tag{69}$$
since H is a group we have
$$\sigma\gamma^{-1} = \begin{pmatrix} 2 & 3 & 1 & 4 \dots & k \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 & \dots & k \end{pmatrix}^{-1} \tag{70}$$
$$= \begin{pmatrix} 2 & 3 & 1 & 4 & \dots & k \end{pmatrix} \begin{pmatrix} k & \dots & 4 & 3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 4 \end{pmatrix} \tag{71}$$
2. Suppose γ has at least two 3-cycles as factors, say $\gamma_1 = (1,2,3), \gamma_2 = (4,5,6)$. Then
$$\sigma = (3,4,5)\gamma(3,4,5)^{-1} = (1,2,4)(3,6,5)\gamma_3 \dots \gamma_m \in H \tag{72}$$
and again we have
$$\sigma\gamma^{-1} = \begin{pmatrix} 1 & 2 & 4 \end{pmatrix} \begin{pmatrix} 3 & 6 & 5 \end{pmatrix} \begin{pmatrix} 4 & 5 & 6 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}^{-1} \tag{73}$$
$$= \begin{pmatrix} 1 & 6 & 3 & 4 & 5 \end{pmatrix} \tag{74}$$
which is a 5-cycle, and we are done by case 1.
3. Suppose γ has precisely one 3-cycle factor and all others are transposisions. If the 3-cycle is $\gamma_1 = (1,2,3)$, then $\gamma^2 = (1,2,3)^2 = (1,3,2)$ is a 3-cycle.
4. Suppose γ is the product of disjoint transpositions. Say $\gamma_1 = (1,2), \gamma_2 = (3,4)$. Then as before
$$\sigma = \begin{pmatrix} 1 & 2 & 4 \end{pmatrix} \gamma \begin{pmatrix} 1 & 2 & 4 \end{pmatrix}^{-1} \implies \sigma\gamma^{-1} = \begin{pmatrix} 1 & 4 \end{pmatrix} \begin{pmatrix} 2 & 3 \end{pmatrix} \in H \tag{75}$$

Since $n \geq 5$ by our theorem hypothesis, the permutation $\tau = (2,3,5) \in A_n$, and so

$$\tau \begin{pmatrix} 1 & 4 \end{pmatrix} \begin{pmatrix} 2 & 3 \end{pmatrix} \tau^{-1} = \begin{pmatrix} 1 & 4 \end{pmatrix} \begin{pmatrix} 3 & 5 \end{pmatrix} \in H \tag{76}$$

$$\implies \begin{pmatrix} 1 & 4 \end{pmatrix} \begin{pmatrix} 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 4 \end{pmatrix} \begin{pmatrix} 3 & 5 \end{pmatrix} = \begin{pmatrix} 2 & 5 & 3 \end{pmatrix} \in H \tag{77}$$

2.3 Quotient Groups

Now that we know about normal subgroups, this allows us to endow on the quotient set a group structure.

Definition 2.6 (Quotient Group)

Given a group G and a normal subgroup H, the **quotient group** G/H is the group of left cosets aH with

1. the operation $(aH) \cdot (bH) \coloneqq (ab)H$
2. the identity element eH.
3. inverses $(aH)^{-1}) = (a^{-1})H$.

and order $|G/H| = |G|/|H|$.

Proof. We verify the properties of a group.

1. Suppose as above that $aH = a'H$ and $bH = b'H$. Then $a' = ah$ and $b' = bk$ for some $h, k \in H$. Since H is normal, $b^{-1}hb = h'$ for some $h' \in H$. Therefore,

$$a'b' = (ah)(bk) = a(hb)k = (abh')k = (ab)(h'k) \in (ab)H \tag{78}$$

   and so $(ab)H = (a'b')H$.
2. eH is indeed the identity since $(aH)(eH) = (ae)H = aH$ and $(eH)(aH) = (ea)H = aH$.
3. Inverses are the same logic.
4. Associativity follows from associativity in G.

Finally, by Lagrange's theorem, the order is as stated.

Since the quotient defines a *congruence* class, this makes it a group homomorphism.

Theorem 2.10 (Quotient Maps are Homomorphisms)

The map $p : G \to G/H$ is a group homomorphism.

Proof. Follows immediately from the definition.

It's a bit hard thinking of an intuitive picture of a normal subgroup. Unless you sit down and try to prove that a subgroup is normal, it's difficult to tell right away. The following lemma characterizes normal subgroups in a different manner.

Lemma 2.11 (Normal Subgroup as Kernel)

A subgroup $H \subset G$ is normal if and only if there exists a group homomorphism $\phi : G \to G'$ with $\ker \phi = H$.

Proof. We prove bidirectionally.

1. ($\to$). Since H is normal, we can form the quotient group G/H. Let $\phi : G \to G/H$ be defined

$\phi(a) = aH$. Then,

$$\ker \phi = \phi^{-1}(eH) = \{a \in G \mid aH = eH = H\} \tag{79}$$
$$= \{a \in G \mid a \in H\} \tag{80}$$

Therefore, ϕ is a homomorphism because $\phi(ab) = abH = (aH)(bH)$.

2. ($\leftarrow$) Assume there is a group homomorphism ϕ. Then, $\ker \phi \subset G$ is a subgroup proven in 1.9. Now consider any $g \in G$. Then

$$\phi(ghg^{-1}) = \phi(g)\phi(h)\phi(g^{-1}) = \phi(g) \cdot e \cdot \phi(g)^{-1} = e \implies ghg^{-1} \in \ker \phi \tag{81}$$

Now that we can construct quotient groups, we would like to see if they are isomorphic to any current groups that we know. More specifically, if we have a normal subgroup $H \subset G$, we can cleverly think of some other group G' and construct a group homomorphism $f : G \to G'$ such that $H = \ker f$. If we can do this, then we can construct a nice isomrophism from G/H to G'. Recall a similar theorem in point set topology: given a topological space $(X, \mathscr{T})$ and its quotient space, if we can construct a map from X to a cleverly chosen space Z that agrees with the quotient, then this induces a homeomorphism $X \cong Z$.

Theorem 2.12 (Fundamental Group Homomorphism Theorem)

Let $f : G \to G'$ be a surjective homomorphism.[a] Then $G/\ker f \simeq G'$.[b]

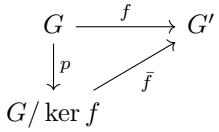


Figure 6: Given f and the projection map $p : G \to G/\ker f$, this induces an isomorphism $\bar{f}$ such that $f = \bar{f} \circ p$.

[a] Sometimes called an *epimorphism*.
[b] Note that if f is not surjective, we can just have it be surjective by restricting G' to be the image of f.

Proof. Let $H = \ker f$, which is then a normal subgroup from 2.11. Now we define a homomorphism

$$\bar{f} : G/H \to G', \qquad \bar{f}(aH) = f(a) \tag{82}$$

We check the following.

1. $\bar{f}$ is well defined. If we have $a, a' \in G$ with $aH = a'H$, then $a' = ah$ for some $h \in H = \ker f$. So $f(a') = f(ah) = f(a)f(h) = f(a)$.
2. $\bar{f}$ is a homomorphism. We see that

$$\bar{f}((aH)(bH)) = \bar{f}((ab)H) \tag{83}$$
$$= f(ab) \tag{84}$$
$$= f(a)f(b) \tag{85}$$
$$= \bar{f}(aH)\bar{f}(bH) \tag{86}$$

3. $\bar{f}$ is surjective. This is trivially true since if not, then $f = \bar{f} \circ p$ cannot be surjective.
4. $\bar{f}$ is injective. By 1.9, it suffices to show that $\ker \bar{f}$ is trivial. Suppose $aH \in \ker \bar{f}$. Then $\bar{f}(aH) = f(a) = e_{G'} \implies a \in H \implies aH = eH$.

Example 2.5 (Cyclic Groups)

$(k\mathbb{Z}, +) \subset (\mathbb{Z}, +)$ is a normal subgroup. Our intuition might tell us that the cosets of the form $k\mathbb{Z}, 1 + k\mathbb{Z}, \dots, (k-1) + k\mathbb{Z}$ behave like integers modulo k, i.e. a cyclic group. Therefore, we can construct the map

$$f : \mathbb{Z} \to Z_k, \qquad f(x) = x \pmod{k} \tag{87}$$

This is a homomorphism and also $\ker f = k\mathbb{Z}$, and so by the fundamental homomorphism theorem

$$\frac{\mathbb{Z}}{k\mathbb{Z}} \simeq Z_k \tag{88}$$

By establishing the connection between the integers and cyclic groups, we establish the notation $Z_k = \mathbb{Z}_k$.

Example 2.6 (Quotient of Reals over Integers)

We can see that $(\mathbb{Z}, +) \subset (\mathbb{R}, +)$ is a normal subgroup. Our intuition might tell us that the cosets (which are disconnected sets consisting of isolated points $\{\dots, x-1, x, x+1, \dots\}$) behave sort of like the rotations on a circle S^1. Therefore, let us construct a map

$$f : \mathbb{R} \to S^1, \qquad f(x) = \cos 2\pi x + i \sin 2\pi x \subset \mathbb{C} \tag{89}$$

Since $f(x+y) = f(x)f(y)$, it follows that f is a homomorphism. On the other hand, $\ker f = \{x \in \mathbb{R} \mid \cos 2\pi x = 1, \sin 2\pi x = 0\} = \mathbb{Z}$. Therefore by the fundamental homomorphism theorem, we have

$$\mathbb{R}/\mathbb{Z} \simeq S^1 \tag{90}$$

Example 2.7 (Determinant)

The determinant $\det : \mathrm{GL}_n(\mathbb{F}) \to \mathbb{F}^* = \mathbb{F} \setminus \{0\}$ is a surjective group homomorphism (under multiplication on $\mathbb{F}$). Therefore,

$$\frac{\mathrm{GL}_n(\mathbb{F})}{\mathrm{SL}_n(\mathbb{F})} \simeq \mathbb{F}^* \tag{91}$$

2.4 Orbits and Stabilizers

Definition 2.7 (Orbits)

Let G be a transformation group on set X. Points $x, y \in X$ are equivalent with respect to G if there exists an element $g \in G$ such that $y = gx$. This has already been defined through the equivalence of figures before. This relation splits X into equivalence classes, called **orbits**. Note that cosets are the equivalence classes of the transformation group G; oribits are those of X. We denote it as

$$Gx \equiv \{gx \mid g \in G\} \tag{92}$$

By definition, transitive transformation groups have only one orbit.

Definition 2.8

The subgroup $G_x \subset G$, where $G_x \equiv \{g \in G | gx = x\}$ is called the **stabilizer** of x.

Example 2.8

The orbits of $O(2)$ are concentric circles around the origin, as well as the origin itself. The stabilizer of 0 is the entire $O(2)$.

Example 2.9

The group S_n is transitive on the set $\{1, 2, ..., n\}$. The stabilizer of $k, (1 \leq k \leq n)$ is the subgroup $H_k \simeq S_{n-1}$, where H_k is the permutation group that does not move k at all.

Theorem 2.13

There exists a 1-to-1 injective correspondence between an orbit G_x and the set G/G_x of cosets, which maps a point $y = gx \in Gx$ to the coset gG_x.

Corollary 2.14

If G is a finite group, then

$$|G| = |G_x||Gx| \tag{93}$$

In fact, there exists a precise relation between the stabilizers of points of the same orbit, regardless of G being finite or infinite:

$$G_{gx} = gG_xg^{-1} \tag{94}$$

2.5 Centralizers and Normalizers

2.6 Lattice of Subgroups

2.7 Exercises

Exercise 2.1 (Shifrin 6.2.2)

Prove that $\mathbb{Z}_7^\times \cong \mathbb{Z}_6$. (It is crucial to remember that we multiply in $\mathbb{Z}_7^\times$ and add in $\mathbb{Z}_6$.)

Solution. Both groups are of order 6, and so $\mathbb{Z}_7^\times$—which is indeed a group (since it is the group of units of the ring $(\mathbb{Z}_7, +, \times)$)—must be isomorphic to either $\mathbb{Z}_6$ or S_3. However, S_3 is not abelian, while $\mathbb{Z}_7^\times$ is, so it must be the case that it is isomorphic to $\mathbb{Z}_6$.

Exercise 2.2 (Shifrin 6.2.15.a/b)

The **dihedral group** of order $2n$, denoted $\mathcal{D}_n$, is given by $\{\rho^i\psi^j : 0 \leq i < n, 0 \leq j \leq 1\}$ subject to the rules $\rho^n = e$, $\psi^2 = e$, and $\psi\rho\psi^{-1} = \rho^{-1}$.

1. Check this is really a group. That is, what is $(\rho^i\psi^j)^{-1}$, and what is the product $(\rho^i\psi^j)(\rho^k\psi^\ell)$?
2. Check that $\mathcal{T} \cong \mathcal{D}_3$ and $S_q \cong \mathcal{D}_4$.

Solution. We check the properties of a group. The following identity is useful:

$$(\psi\rho\psi^{-1})^{n-i} = (\rho^{-1})^{n-i} \implies \psi\rho^{n-i}\psi^{-1} = \rho^i \implies \psi\rho^{n-i} = \rho^i\psi \tag{95}$$

1. *Closure.* From simplifying according to the first two rules, we will automatically adjust the exponents to be $i, k < n$ (by subtracting out multiples of n) and $j \in \{0, 1\}$ (by subtracting out

multiples of 2). Going case by case,
   (a) $j = 0, l = 0$. $\rho^i \rho^k = \rho^{i+k}$.
   (b) $j = 0, l = 1$. $\rho^i \rho^k \psi = \rho^{i+k} \psi$.
   (c) $j = 1, l = 0$. $\rho^i \psi \rho^k = \rho^i \rho^{n-k} \psi = \rho^{n-k+i} \psi$.
   (d) $j = 1, l = 1$. $\rho^i \psi \rho^k \psi = \rho^i \psi \psi \rho^{n-k} = \rho^i \rho^{n-k} = \rho^{n-k+i}$.
2. *Identity.* The identity is $e = \rho^0 \psi^0$. We can see that $e \rho^i \psi^j = \rho^i \psi^j e = \rho^{i+0} \psi^j$.
3. *Inverse.* We have $\psi \rho \psi^{-1} = \psi \rho \psi = \rho^{-1} \implies \psi \rho = \rho^{-1} \psi^{-1} = (\psi \rho)^{-1}$. Therefore,

$$(\rho^i \psi^j)^{-1} = \begin{cases} \rho^{n-i} & \text{if } j = 0 \\ \rho^i \psi & \text{if } j = 1 \end{cases} \tag{96}$$

   which are both of the correct form and therefore in $\mathcal{D}_n$. To verify, we see that $\rho^i \rho^{n-i} = \rho^n = e$, and $(\rho^i \psi)(\rho^i \psi) = \rho^i \psi \psi \rho^{n-i} = \rho^i \rho^{n-i} = e$.
4. *Associativity.* Can also be proven tediously but problem only asked to state the product and inverse.

For (b) for $\mathcal{T}$, we can explicitly look at the multiplication tables and see that they are isomorphic. We denote r_1, r_2 as the 120 and 240 degree rotations, and f_1, f_2, f_3 as the flips across each axis.

	e	ρ	ρ^2	ψ	$\rho\psi$	$\rho^2\psi$
e	e	ρ	ρ^2	ψ	$\rho\psi$	$\rho^2\psi$
ρ	ρ	ρ^2	e	$\rho^2\psi$	ψ	$\rho\psi$
ρ^2	ρ^2	e	ρ	$\rho\psi$	$\rho^2\psi$	ψ
ψ	ψ	$\rho^2\psi$	$\rho\psi$	e	ρ^2	ρ
$\rho\psi$	$\rho\psi$	ψ	$\rho^2\psi$	ρ	e	ρ^2
$\rho^2\psi$	$\rho^2\psi$	$\rho\psi$	ψ	ρ^2	ρ	e

(a) $\mathcal{D}_3$

	e	r_1	r_2	f_1	f_2	f_3
e	e	r_1	r_2	f_1	f_2	f_3
r_1	r_1	r_2	e	f_3	f_1	f_2
r_2	r_2	e	r_1	f_2	f_3	f_1
f_1	f_1	f_2	f_3	e	r_2	r_1
f_2	f_2	f_3	f_1	r_1	e	r_2
f_3	f_3	f_1	f_2	r_2	r_1	e

(b) $\mathcal{T}$

For S_q, it is tedious to write the full table, so we construct the isormorphisms using the generators. For S_q, the symmetry group of the square consists of 8 elements: the 4 rotations r_1, r_2, r_3, r_4 (of 90, 180, 270, and 360=0 degrees), and the flips f_1, f_2, f_3, f_4 (across each axis). Now we construct the function $g : \mathcal{D}_3 \to \mathcal{T}$ such that $f(\rho) = r_1$ and $f(\psi) = f_1$. Then we can see that

$$g(\rho^4) = g(e) = e = r_1^4 = g(\rho^4), \qquad g(\psi^2) = g(e) = e = f_1^2 = g(\psi)^2 \tag{97}$$

since 90 degrees rotated 4 times is 0 degrees, the identity, and two flips across the same axis is also the identity. Finally, we have

$$g(\psi\rho\psi) = g(\rho^{-1}) = r_1^{-1} = r_3 = f_1 r_1 f_1 = g(\psi) g(\rho) g(\psi) \tag{98}$$

Where $r_1^{-1} = r_3$ since a rotation of 270 after a 90 is the same as rotation by 360=0, and $r_3 = f_1 r_1 f_1$ is the change of basis symmetry observed in Shifrin Example 6.1.5. Therefore the rules match, making it a homomorphism, and since the order is the same ($\mathcal{D}_3$ has $4 \times 2 = 8$ elements from looking at the indices), this is an isomorphism.

Exercise 2.3 (Shifrin 6.3.8)

Let $H \subset G$ be a subgroup, and let $a \in G$ be given. Prove that $aHa^{-1} \subset G$ is a subgroup (called a **conjugate subgroup** of H). Prove, moreover, that it is isomorphic to H (cf. Exercise 6.2.12).

Solution. Let $x, y \in aHa^{-1}$. Then $x = ah_x a^{-1}, y = ah_y a^{-1}$ for some $h_x, h_y \in H$. Therefore,
1. It is closed. $xy = (ah_x a^{-1})(ah_y a^{-1}) = ah_x(a^{-1}a)h_y a^{-1} = ah_x h_y a^{-1} \in aHa^{-1}$ since $h_x h_y \in H$

by closure.
2. It has an identity since $e \in H \implies aea^{-1} = aa^{-1} = e \in aHa^{-1}$.
3. It has inverses since given $x \in H$ as above with inverses x^{-1}, we see that $(axa^{-1})^{-1} = (a^{-1})^{-1}x^{-1}a^{-1} = ax^{-1}a^{-1} \in aHa^{-1}$ since $x^{-1} \in H$ by H being a group.
4. Associativity is inherited from G.

It suffices to show that this is injective, since the map $\iota : H \to aHa^{-1}$ is surjective by definition. Given $x, y \in aHa^{-1}$ with $x = y$, we have $ah_xa^{-1} = ah_ya^{-1}$, and multiplying by a on the right and then a^{-1} on the left, we get $h_x = h_y$.

Exercise 2.4 (Shifrin 6.3.11)

Prove that a group of order n has a proper subgroup if and only if n is composite.

Solution. We prove bidirectionally. Call the group G and subgroup H.
1. ($\rightarrow$). Assume n is prime. Then by Lagrange's theorem $|H|$ must divide n, and so $|H| = 1$ or n, neither of which results in a proper subgroup.
2. ($\leftarrow$). Assume G has a proper subgroup H. Since it is proper, $|H| \neq 1, n$. Then by Lagrange's theorem, $|H|$ divides n, which implies that n is composite.

Exercise 2.5 (Shifrin 6.3.13)

Suppose $H, K \subset G$ are subgroups of orders 5 and 8, respectively. Prove that $H \cap K = \{e\}$.

Solution. Let us take an arbitrary element in $x \in H \cap K$ and consider the cyclic group $\langle x \rangle$. By Lagrange's Theorem, the order $|x|$ in H must be either 1 or 5, while the order in K must be $1, 2, 4, 8$. Therefore, $|x| = 1$ and so $x = e$.

Exercise 2.6 (Shifrin 6.3.17)

1. Prove that a group G of even order has an element of order 2. (Hint: If $a \neq e$, a has order 2 if and only if $a = a^{-1}$.)
2. Suppose m is odd, $|G| = 2m$, and G is abelian. Prove G has precisely one element of order 2. (Hint: If there were two, they would provide a Klein four-group.)
3. Prove that if G has exactly one element of order 2, then it must be in the center of G.

Solution. Listed.
1. Assume the contrary and take $H = G \setminus \{e\}$. Then $|H|$ is odd, and since no element has order 2, every element must be associated with a unique inverse a, a^{-1}. But this cannot happen since $|H|$ is odd. Therefore there must be at least one element of order 2.
2. It has at least 1 element of order 2 from (1). Now assume that there are two, call them a, b. Then $ab \neq a, b$ and ab also has order 2 since $(ab)(ab) = abba = aa = e$. Therefore, calling $c = ab$, we have $ac = ca = aab = b$ and $bc = cb = abb = a$. This fully defines the multiplication table for the Klein 4 group K of order 4. Therefore, by Lagrange's theorem, we have found a subgroup K and so $|K|$ must divide G. However, this would mean that m must be even, a contradiction. Therefore there is only one such unique a.
3. Given $a \in G$ with $|a| = 2$, we wish to show that it is an element of $Z = \{b \in G \mid bx = xb \forall x \in G\}$.[a] Consider $z = x^{-1}ax$. We have

$$z^2 = (x^{-1}ax)^2 = x^{-1}axx^{-1}ax = x^{-1}a^2x = x^{-1}x = e \tag{99}$$

which means that z also has order 2. But since this is unique, it must be that $z = a$. Therefore, by multiplying x on the left, we get

$$x^{-1}ax = a \implies ax = xa \tag{100}$$

[a] I am using the definition of center defined in Shifrin 6.3.7.

Exercise 2.7 (Assigned)

Find all group homomorphisms $\mathbb{Z}_n \to \mathbb{Z}_m$. (Your answer will depend on n and m.)

Solution. Given a homomorphism, f, we must have $f(0) = 0$. Let $f(1) = k$. Note that the value of $f(1) = k$ completely determines the homomorphism since the image of every other $l \in \mathbb{Z}_n$ is defined by

$$f(l) = f(\underbrace{1 + \ldots + 1}_{l \text{ times}}) = \underbrace{k + \ldots + k}_{l \text{ times}} \tag{101}$$

Since the image of f must be a cyclic subgroup of $\mathbb{Z}_m$, we must satisfy

$$0 = f(0) = f(\underbrace{1 + \ldots + 1}_{n \text{ times}}) \tag{102}$$

$$= \underbrace{k + \ldots + k}_{n \text{ times}} \tag{103}$$

and so $m \mid nk$. Therefore, k must be a multiple of $m/\gcd(n, m)$. So all homomorphisms are determined by the set

$$\left\{ k = \frac{am}{\gcd(n, m)} \;\middle|\; a \in \mathbb{N}, 0 \leq k \leq m - 1 \right\} \tag{104}$$

which we can see ranges from $0 \leq a < \gcd(n, m)$, and so the total number of homomorphisms is $\gcd(n, m)$. Note that there is always the trivial homomorphism when $a = 0$, i.e. everything maps to 0. For example, if we have $f : \mathbb{Z}_{14} \to \mathbb{Z}_{21}$, we have $k = 0, 3, 6, 9, 12, 15, 18$.

3 Group Actions

3.1 Sylow Theorems

3.2 Exercises

4 Classification of Groups

4.1 Direct Products

Definition 4.1 (Direct Product)

The **direct product** of two groups $(G, \cdot)$ and $(H, *)$ is the set

$$G \times H \equiv \{(g, h) \mid g \in G, h \in H\} \tag{105}$$

equipped with the operation

$$(g_1, h_1) \cdot (g_2, h_2) \coloneqq (g_1 \cdot g_2, h_1 * h_2) \tag{106}$$

Proof. It is pretty trivial to see that this is a group.

Example 4.1 (General Affine Group)

The **general affine group** is defined

$$\mathrm{GA}(V) \equiv \mathrm{Tran}(V) \times \mathrm{GL}(V) \tag{107}$$

Example 4.2 (Galileo Group)

The **Galileo Group** is the transformation group of spacetime symmetries that are used to transform between two reference frames which differ only by constant relative motion within the constructs of Newtonian physics. It is denoted

$$\mathrm{Tran}\,\mathbb{R}^4 \times H \times \mathrm{O}(3) \tag{108}$$

where H is the group of transformations of the form

$$(x, y, z, t) \mapsto (x + at, y + bt, z + ct, t) \tag{109}$$

Example 4.3 (Poincaré Group)

The **Poincaré Group** is the symmetry group of spacetime within the principles of relativistic mechanics, denoted

$$G = \mathrm{Tran}\,\mathbb{R}^4 \times \mathrm{O}_{3,1} \tag{110}$$

where $\mathrm{O}_{3,1}$ is the group of linear transformations preserving the polynomial

$$x^2 + y^2 + z^2 - t^2 \tag{111}$$

4.2 Semidirect Products

4.3 Classification of Finite Abelian Groups

Theorem 4.1 (Groups of Order 1, 2, 3)

We have the following.

1. There is only one group of order 1.

$$Z_1 \simeq S_1 \simeq A_2 \tag{112}$$

2. There is only one group of order 2.
$$Z_2 \simeq S_2 \simeq D_2 \tag{113}$$
3. There is only one group of order 3.
$$Z_3 = A_3 \tag{114}$$

Theorem 4.2 (Groups of Order 4)

There are two groups of order 4.
$$Z_4, \qquad Z_2^2 \simeq D_4 \tag{115}$$

4.4 Group Extensions

4.5 Classification of Simple Groups of Small Order

Theorem 4.3 (Classification of Simple Groups of Small Order)

The following are the only groups of order n. You can notice that it is dominated by direct products of cyclic groups, since they exist for every order, while the other types increase in order very fast.

n	**Abelian Groups**	**Non-Abelian Groups**
1	$\{e\}$ (trivial group)	None
2	$\mathbb{Z}_2 = S_2 = \text{Dih}(1)$	None
3	$\mathbb{Z}_3 = A_3$	None
4	$\mathbb{Z}_4$, $\mathbb{Z}_2 \times \mathbb{Z}_2 = \text{Dih}(2)$	None
5	$\mathbb{Z}_5$	None
6	$\mathbb{Z}_6 = \mathbb{Z}_3 \times \mathbb{Z}_2$	$S_3 = \text{Dih}(3)$
7	$\mathbb{Z}_7$	None
8	$\mathbb{Z}_8$, $\mathbb{Z}_4 \times \mathbb{Z}_2$, $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$	$D_4 = \text{Dih}(4)$, Q_8 (quaternion)
9	$\mathbb{Z}_9$, $\mathbb{Z}_3 \times \mathbb{Z}_3$	None
10	$\mathbb{Z}_{10} = \mathbb{Z}_5 \times \mathbb{Z}_2$	$D_5 = \text{Dih}(5)$
11	$\mathbb{Z}_{11}$	None
12	$\mathbb{Z}_{12} = \mathbb{Z}_4 \times \mathbb{Z}_3$, $\mathbb{Z}_6 \times \mathbb{Z}_2$, $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3$	A_4, $D_6 = \text{Dih}(6)$, $\mathbb{Z}_3 \rtimes \mathbb{Z}_4$ (dicyclic)

Figure 8: Classification of groups up to order 12.

4.6 Exercises

5 Rings

We have extensively talked about groups, and now we look at an algebraic structure called a ring that has two operations. As we introduce rings, we will use the integers as the primary structure to demonstrate our theorems, along with the ring of continuous functions and the ring of matrices.

Definition 5.1 (Ring)

A **ring** is a set $(R, +, \times)$ equipped with two operations, called addition and multiplication. It has properties:

1. R is an abelian group with respect to $+$, where we denote the additive identity as 0 and the additive inverse of x as $-x$.
2. R is a monoid with respect to $\times$, where we denote the multiplicative identity as 1, also known as the **unity**.
3. $\times$ is both left and right distributive with respect to addition $+$

$$a \times (b + c) = a \times b + a \times c \tag{116}$$
$$(a + b) \times c = a \times c + b \times c \tag{117}$$

for all $a, b, c \in \mathbb{R}$.

If $\times$ is commutative, R is called a **commutative ring**.

In fact, in some cases associativity (in multiplication) or the existence of the multiplicative identity is not even assumed, though we will do it here.[2] It turns out that the existence of a multiplicative inverse, also called a *unity*, forces addition to be abelian. Try computing the product $(1+1)(a+b)$ in two different ways.

$$(1 + 1)(a + b) = 1(a + b) + 1(a + b) = a + b + a + b \tag{118}$$
$$(1 + 1)(a + b) = (1 + 1)a + (1 + 1)b = a + a + b + b \tag{119}$$

and so from the group properties, we necessarily have $b + a = a + b$.

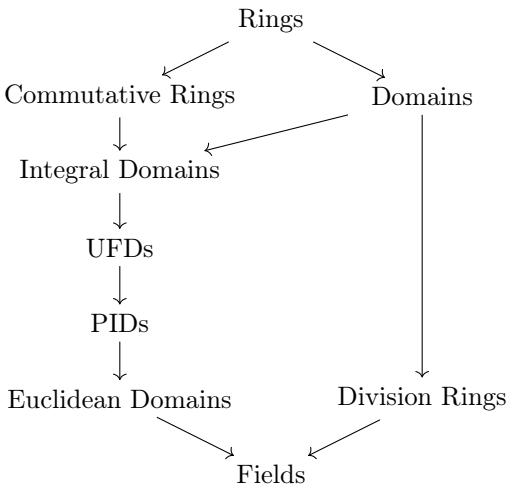


Figure 9: Basic hierarchy of rings with UFDs included.

Since a ring is a group with respect to addition, we know from 1.3 that additive inverses are unique. However, we can say a little more with rings because of the distributive property.

[2] If a multiplicative identity is not assumed, then this is called an *rng*, or a *rung*.

Lemma 5.1 (Additive Inverses)

Let R be a ring. Then, for all $a, b \in R$,
1. $0a = a0 = 0$.
2. $(-a)b = a(-b) = -(ab)$.
3. $(-a)(-b) = ab$.
4. The identity 1 is unique and $-a = (-1)a$.

Proof. We can see that

$$-1 + 1 = 0 \implies (-1 + 1) \times a = 0 \times a \tag{120}$$
$$\tag{121}$$

and therefore by definition $-1 \times a$ must be the additive inverse.

We will provide some examples of rings, though we have not properly defined some of them yet. We have only properly defined the power set ring and the integer ring in set theory. As for the rest, I hope the reader is familiar enough with these materials to at least recognize that they are rings.

Example 5.1 (Power Set)

Given a set X, $(2^X, \triangle, \cap)$ is a commutative associative ring with respect to the operations of symmetric difference $M \triangle N \coloneqq (M \backslash N) \cup (N \backslash M)$ and intersection. The additive identity is $\emptyset$ and the multiplicative identity is X. We can clearly see that both operations are commutative and $\cap$ is associative.

$$M \triangle N = (M \setminus N) \cup (N \setminus M) \equiv N \triangle M$$
$$M \cap N = N \cap M$$
$$M \cap N \cap P = (M \cap N) \cap P = M \cap (N \cap P)$$

Example 5.2 (Number Systems)

$(\mathbb{Z}, +, \times)$, $(\mathbb{Q}, +, \times)$, $(\mathbb{R}, +, \times)$, $(\mathbb{C}, +, \times)$ are all commutative rings, with additive and multiplicative identities 0 and 1.

Example 5.3 (Matrix Rings)

The set of matrices $\mathbb{R}^{n \times n}$[a] forms a noncommutative ring under matrix addition $+$ and multiplication $\times$. It has the additive and multiplicative identities 0 and I_n. This forms a non-commutative ring for $n > 1$, even when R is commutative.

[a] really over any field and even more generally a ring R

Example 5.4 (Polynomials)

The set of all polynomials over the reals are a ring.

Example 5.5 (Continuous Functions)

The set of all continuous functions $f : \mathbb{R} \to \mathbb{R}$ is a ring under point-wise addition and multiplication.

Next, just like how we did for groups, we can talk about subrings.

Definition 5.2 (Subring)

Given ring $(R, +, \times)$ a **subring** $(S, +, \times)$ is a ring such that $S \subset R$. S is called a **proper subring** if $S \subsetneq R$.

Example 5.6 (Quadratic Field)

Let $D \in \mathbb{Q}$ be a square-free rational number (as in D cannot be expressed as the perfect square of another rational). Now let us consider the set

$$\mathbb{Q}(\sqrt{D}) \coloneqq \{a + b\sqrt{D} \mid a, b \in \mathbb{Q}\} \tag{122}$$

equipped with component-wise addition and multiplication where $\sqrt{D}^2 = D$ and the rational commute with $\sqrt{D}$. This is indeed a commutative subring of $\mathbb{R}$.

Example 5.7 (Gaussian Integers)

The **Gaussian integers** is the set

$$\mathbb{Z}[i] \coloneqq \{a + bi \mid a, b \in \mathbb{Z}\} \subset \mathbb{C} \tag{123}$$

equipped with component-wise addition and multiplication where $i^2 = 1$ and reals commute with i. This is indeed a commutative subring of $\mathbb{C}$.

Theorem 5.2 (Intersections of Subrings is a Subring)

If S_1, S_2 are subrings of R, then $S_1 \cap S_2$ is a subring.

Finally, we will mention a product ring.

Definition 5.3 (Direct Product of Rings)

Given rings $(R, +_R, \times_R)$ and $(S, +_S, \times_S)$, the direct product of the rings is the set $R \times S$ with the operations

1. $(r_1, s_1) + (r_2, s_2) \coloneqq (r_1 +_R r_2, s_1 +_S s_2)$.
2. $(r_1, s_1) \times (r_2, s_2) \coloneqq (r_1 \times_R r_2, s_1 \times_S s_2)$.

Proof. The proof is standard.

5.1 Ring Homomorphisms

So far, we have talked about many properties of rings but have not thoroughly gone over their classification. This is what we will do in this section, just like how we have classified groups. It turns out that classifying rings is significantly harder to do so, so we will talk about some low-order finite rings and provide some examples of isomorphisms between more complex rings.

Definition 5.4 (Ring Homomorphism, Isomorphism)

A **ring homomorphism** $f : R \to S$ is a function that satisfies for all $a, b \in R$
1. $f(a+b) = f(a) + f(b)$
2. $f(ab) = f(a)f(b)$
3. $f(1_R) = 1_S$[a]

for all $a, b \in R$.[b] Furthermore,
1. A **ring isomorphism** is a bijective ring homomorphism, and we call rings R and S isomorphic, denoted $R \simeq S$ if there exists an isomorphism between them.
2. A **ring endomorphism** is a ring homomorphism onto itself.
3. A **ring automorphism** is an isomorphism from a ring to itself.

[a] The reason we need this third is that while f is a group homomorphism with respect to $+$, it automatically follows that $f(0) = 0$. However f is only a monoid homomorphism w.r.t. $\times$, and so we need this extra constraint.

[b] Note that the first is equivalent to it being a group homomorphism between $(R, +)$ and $(S, +)$. The second property may look like it is a group homomorphism between $(R, \times)$ and $(S, \times)$, but remember that neither are groups and it just states that closure distributes. Combined with the fact that the multiplicative identity matches, f is really a homomorphism of *monoids*.

Example 5.8 (Homomorphisms of Rings)

We provide some simple examples of ring homomorphisms.
1. The identity map $\iota : R \to R$ is a ring automorphism.
2. If $R \subset S$ as rings, then the canonical injection map $\iota : R \to S$ is a ring homomorphism.
3. Complex conjugation $z \in \mathbb{C} \mapsto \bar{z} \in \mathbb{C}$ is a ring automorphism.
4. Differentiation is a ring automorphism over the polynomial ring $\mathbb{R}[x]$.

Definition 5.5 (Kernel)

The **kernel** of a ring homomorphism $f : R \to S$ is the preimage of $0 \in S$.[a]

[a] Note that this is the additive identity, not the multiplicative identity. We must specify which identity, unlike a group which has just one identity.

Lemma 5.3 (Images and Kernels of Ring Homomorphisms)

If $f : R \to S$ is a ring homomorphism, then
1. $\text{Im}\, f$ is a subring of S.
2. f is injective iff $\ker f = \{0\}$.

Proof. For the first claim, let $x, y, z \in \text{Im}\, f$. Then $x = f(a), y = f(b), z = f(c)$ for some $a, b \in R$.
1. *Closed under Addition.* $x + y = f(a) + f(b) = f(a + b) \in \text{Im}\, f$.
2. *Associative under Addition.* $(x + y) + z = f(a + b) + f(c) = f((a + b) + c) = f(a + (b + c)) = f(a) + f(b + c) = x + (y + z)$
3. *Additive Identity.* $f(0) = 0$
4. *Additive Inverses.* We claim that $x^{-1} = f(a)^{-1} = f(a^{-1})$. Indeed, we have $f(a)f(a^{-1}) = f(aa^{-1}) = f(1) = 1$.
5. *Closed under Multiplication.* $xy = f(a)f(b) = f(ab) \in \text{Im}\, f$.
6. *Multiplicative Identity.* $f(1) = 1$.

For the second claim, we prove bidirectionally.
1. $(\to)$. Let $\ker f \neq \{0\}$ and call its nonzero element k. Then, $f(a + k) = f(a) + f(k) = f(a) + 0 = f(a)$, and so $f(a) = f(a + k)$, which means f is not injective.

2. ($\leftarrow$). Assume that f is not injective. Then there exists $a, b \in R$ s.t. $f(a) = f(b)$. This means that $0 = f(a) - f(b) = f(a - b)$, and so $a - b \in \ker f$.

Note that $\ker f$ is *not* a subring, and we can quickly verify this by noticing that the identity element does not necessarily have to be in the kernel. However, we will see later that this is a specific instance of a more general structure called an *ideal*.

Theorem 5.4 (Compositions of Ring Homomorphisms)

Compositions of ring homomorphisms are ring homomorphisms.

Proof. Let $R \xrightarrow{f} S \xrightarrow{g} T$ be two ring homomorphisms. We can see that
1. $(g \circ f)(a + b) = g(f(a) + f(b)) = g(f(a)) + g(f(b))$.
2. $g(f(ab)) = g(f(a)f(b)) = g(f(a)) + g(f(b))$
3. $g(f(1_R)) = g(1_S) = 1_T$

Now let's focus a bit more on ring isomorphisms. The following should be intuitive.

Lemma 5.5 (Inverse of Ring Isomorphism is an Isomorphism)

If $f : R \to S$ is a ring isomorphism, then f^{-1} is a ring isomorphism.

Proof. Since f is a bijection, f^{-1} is well defined and is a bijection. Now let $x, y \in S$, which implies that $x = f(a), y = f(b)$ for a unique $a, b \in R$. Now we see that f^{-1} satisfies the 3 properties of a ring homomorphism.
1. $f^{-1}(x + y) = f^{-1}(f(a) + f(b)) = f^{-1}(f(a + b)) = a + b = f^{-1}(a) + f^{-1}(b)$.
2. $f^{-1}(xy) = f^{-1}(f(a)f(b)) = f^{-1}(f(ab)) = ab = f^{-1}(a)f^{-1}(b)$.
3. $f^{-1}(1_S) = 1_R$.

Therefore, as a bijective ring homomorphism f^{-1} is also a ring isomorphism.

5.2 Characteristics

Note that given a ring R, we can pay attention to the subring $\langle 1 \rangle$. This must either be isomorphic to $\mathbb{Z}$ or $\mathbb{Z}_n$, so we can think of it being embedded in R.

Theorem 5.6 (Integer Ring Exists in Any Ring)

For every ring R, there exists a unique ring homomorphism $f : \mathbb{Z} \to R$.

Proof. We know that $f(1_\mathbb{Z}) = 1_R$, and so for $n > 0$,

$$\begin{aligned} f(n_\mathbb{Z}) &= f(1_\mathbb{Z} + \ldots + 1_\mathbb{Z}) && (124)\\ &= f(1_\mathbb{Z}) + \ldots + f(1_\mathbb{Z}) && (125)\\ &= 1_R + \ldots + 1_R && (126)\\ &= n_R && (127) \end{aligned}$$

Similarly, we have

$$\begin{aligned} f(-n_{\mathbb{Z}}) &= f(-1_{\mathbb{Z}} - \ldots - 1_{\mathbb{Z}}) && (128)\\ &= f(1_{\mathbb{Z}}) - \ldots - f(1_{\mathbb{Z}}) && (129)\\ &= -1_R - \ldots - 1_R && (130)\\ &= -n_R && (131) \end{aligned}$$

Since $\mathbb{Z}$ is a PID, $\ker f$—which is an ideal—must be principal, and so $\ker f = \langle m \rangle$ for some $m \in \mathbb{Z}$.

Therefore, this motivates the following attribute of a ring, i.e. the smallest $\langle m \rangle$ that embeds (an injective homomorphism) into the ring.

Definition 5.6 (Characteristic Number)

The **characteristic** of ring R, denoted $\text{char}(R)$, is defined equivalently.

1. It is the smallest number of times one must successively add the multiplicative identity 1 to get the additive identity 0.
$$1 + 1 + ... + 1 = 0 \tag{132}$$
If no such number n exists, then $\text{char}(R) = 0$.
2. It is equal to m, where $\ker f = \langle m \rangle$ for the homomorphism defined above.[a]

[a] Note that m always exists since $\mathbb{Z}$ is a PID.

Often, it is not obvious whether two given rings R and S are isomorphic. The characteristic number is preserved across ring isomorphisms and therefore is a good sanity check.

Theorem 5.7 (Preservation of Characteristic Number in a Ring Homomorphism)

$R \simeq S \implies \text{char}(R) = \text{char}(S)$.

Proof.

However, the converse is not true! If so, we would have completely classified all rings just based on their characteristic number, and the study of rings would end pretty soon.

Example 5.9 (Same Characteristic does not Imply Isomorphic)

There exists no isomorphism from $\mathbb{Z}$ to $\mathbb{R}$.

Theorem 5.8 (Wilson's Theorem)

Let $p \in \mathbb{N}$ be prime. Then

$$(p-1)! \equiv -1 \pmod{p} \tag{133}$$

5.3 Divisors and Reducibility

Note that we do not assume that there exists multiplicative inverses in a ring. However, there may be some elements for which multiplicative inverses do exist, i.e. $a, b \in R$ where $ab = 1$.

Definition 5.7 (Unit)

A **unit** of a ring R is an element $u \in R$ that has a multiplicative inverse in R. That is, there exists a $v \in R$ s.t. $uv = vu = 1$.

Example 5.10 (Units and Non-Units)

A ring R may have either none (except for 0 and 1), some, or all of its elements as units.

1. $\mathbb{Z}$ has no non-unity element that is a unit. For example, given $z \in \mathbb{Z}$, there is no element of the form $1/z$.
2. In $\mathbb{Q}$, every nonzero element is a unit. Given any element of form $\frac{p}{q} \in \mathbb{Q}$, the element $\frac{q}{p}$ is the multiplicative inverse.
3. In $\mathbb{Z}_8$, the units are $1, 3, 5, 7$ as $1 \cdot 1 = 3 \cdot 3 = 5 \cdot 5 = 7 \cdot 7 = 1$. However, $0, 2, 4, 6$ are not units since you can look in $\mod 2$.

Another property that we would desire is some sort of decomposition of ring elements as other ring elements. More specifically, the existence of elements a, b such that $ab = 0$ will be of particular interest to us.

Definition 5.8 (Left, Right Divisor)

Let $a, b, r \in R$ a ring.

1. If $ab = r$, then a is said to be a **left divisor** of r and b a **right divisor** of r.
2. a is said to be a left divisor of r if it is a left divisor and a right divisor of r: $ax = ya = r$, but x does not necessarily equal y.
3. If $ab = 0$, then a and b are said to be a **left zero divisor** and **right zero divisor**, respectively.

If R is commutative, then we just call a a **divisor** of r or a **zero divisor**.[a]

[a] a is a right divisor of $b \iff \exists x(xa = b) \iff \exists x(ax = b) \iff a$ is a left divisor.

Definition 5.9 (Reducibility of Elements)

Let R be a commutative ring and $r \in R$ be nonzero and not a unit.

1. Then r is **irreducible** in R if whenever $r = ab$ with $a, b \in R$, at least one of a, b is a unit in R.
2. Otherwise, r is said to be **reducible**, and a, b are said to be **factors**

Note that reducibility is slightly different than the existence of divisors of an element. r may have divisors, but they could also be units. Therefore, irreducibility is weaker than having no divisors. This should not be mixed up with prime/composite elements either, which we have not defined

Lemma 5.9 (Units and Zero Divisors are Mutually Exclusive)

An element $a \in R$ can never be both a unit and a zero divisor.

Proof. Let $a \in R$ be a unit. Then $1 = ab$ for some $b \in R$. Now if a was a zero divisor, then $ra = 0$ for some $r \in R$. However, we have $r(ab) = r1 = r \neq 0 = 0b = (ra)b$, which contradicts associativity.

Let's go through some examples to see that we can have widely differing behavior in terms of divisors.

Example 5.11 (Left Divisor But Not Right Divisor in Matrix Ring)

Let us define the ring

$$R = \begin{pmatrix} \mathbb{Z} & \mathbb{Z}_2 \\ 0 & \mathbb{Z} \end{pmatrix} = \left\{ \begin{pmatrix} x & y \\ & z \end{pmatrix} \;\middle|\; x, z \in \mathbb{Z}, y \in \mathbb{Z}_2 \right\} \tag{134}$$

This should be checked that it is a ring. Let

$$a = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix},\; b = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{135}$$

Then, we can see that $ab = 0$ but $ba \neq 0$. That is, b is a right zero divisor of a, but b is not a left zero divisor.

Note that 0 is divisible by every element since $0a = 0$ for all $a \in R$. Furthermore, if p was a unit, then it can always be multiplied by p^{-1} and then by any $a \in R$ to get the factorization $a = (ap^{-1})p$. So p as a unit would divide *every* element in R.

Example 5.12 (Prime and Composite Elements)

An element in a ring R may either be prime, composite, or neither.

1. In $\mathbb{Z}$, 2 is prime but $6 = 2 \cdot 3$.
2. In $\mathbb{Q}$, there are no such things are prime or composite elements since every nonzero element is a unit.
3. In $\mathbb{Z}_5$, every nonzero element is a unit since $2 \cdot 3 = 4 \cdot 4 = 1$, so calling the elements prime or composite does not make sense.
4. In $\mathbb{Z}_6$, $2, 4$ are not units, so they must be either prime or composite. It turns out that 2 is prime and $4 = 2 \cdot 2$ is composite.

Therefore, in a general ring there is too little structure to determine much about the divisibility of elements. The following lemma is all we have.

Lemma 5.10 (Divisibility of Linear Combinations of Rings Elements)

Let R be a commutative ring and $a, b, d \in R$. If $d \mid a$ and $d \mid b$, then $d \mid (ma + nb)$ for any $m, n \in R$.

One may also intuit that when $a \mid b$, a must be "less than" b. First, in a ring without an order or a norm, this statement doesn't really make sense, and you can indeed have two distinct elements that are divisors of each other!

Example 5.13 (Two Distinct Elements are Divisors of Each Other)

In $\mathbb{Z}_{12}$, we have

1. $2 \mid 10$ since $10 = 2 \cdot 5$, and $10 \mid 2$ since $2 = 10 \cdot 5$.
2. $3 \mid 9$ since $9 = 3 \cdot 3$ and $3 \mid 9$ since $3 = 9 \cdot 7$.

Therefore, this motivates the following definition.

Definition 5.10 (Associate Elements)

Elements a and b are **associated**, denoted $a \sim b$ if either of the following equivalent conditions holds

1. $a \mid b$ and $b \mid a$.

2. $a = ub$, for some unit $u \in R$.

This forms an equivalence relation.

Proof. We first prove the equivalence.

1. Assume that $a \mid b$ and $b \mid a$. Then $a \mid b \implies b = ra$ for some $r \in R$. Furthermore, $b \mid a \implies a = qb$ for some $q \in R$. Therefore, we have $b = ra = (rq)b$, and so $rq = 1$, which means r, q must be units.
2. Assume that $a = ub$ for some unit $u \in R$. Then, $b \mid a$. Now, we can multiply by u^{-1} to get $u^{-1}a = b \implies a \mid b$.

To prove that this is an equivalence relation, we prove the following properties.

1. *Reflexive.* $a \sim a$ since $a = 1a$.
2. *Symmetric.* If $a \sim b$, then $a \mid b$ and $b \mid a$, implying that $b \sim a$.
3. *Transitive.* If $a \sim b, b \sim c$, then there exists units $u, v \in R$ s.t. $a = ub, b = vc$, and so $a = (uv)c$, where uv is a unit with $(uv)^{-1} = v^{-1}u^{-1}$.

Example 5.14 (Associate Elements in a Commutative Ring)

We present some examples of associate elements.

1. In $\mathbb{Z}$, ± 6 are associate elements since $6 = -1 \cdot 6$, where -1 is a unit.
2. In $\mathbb{Z}[i]$, $1 + i$, $1 - i$ are associate elements since $1 + i = i(1 - i)$, where i is a unit since $i \cdot -i = 1$.
3. As above, in $\mathbb{Z}_{12}$ the elements $2, 10$ are associate elements since $2 = 5 \cdot 10$, where 5 is a unit since $5 \cdot 5 = 1$.

Remember that for commutative rings, distinguishing left and right divisors are meaningless, and so we can talk about just *divisors*. Almost all rings that we will deal with are commutative, so let's try to find some properties of commutative rings.

Definition 5.11 (Greatest Common Divisor)

Let R be a commutative ring and let $a, b \in R$ with $b \neq 0$. The **greatest common divisor** of elements a and b—denoted $\gcd(a, b)$—is an element $d \in R$ satisfying:

1. $d \mid a$ and $d \mid b$
2. if $k \mid a$ and $k \mid b$, then $k \mid d$.

If $\gcd(a, b) = 1$, then a and b are said to be **relatively prime**.

Note that in an arbitrary commutative ring, the gcd of two elements always exists since we can at least identify 1, but there may not be a *unique* gcd.

5.4 Ideals

Now assuming that R and S are commutative rings, let's consider a special sort of subset of a commutative ring. Consider the kernel of the ring homomorphism. We can see that if $a, b \in \ker(f)$, then $f(a + b) = f(a) + f(b) = 0 + 0 = 0$, and so $\ker(f)$ is closed under addition. Furthermore, $a \in \ker(f)$ and *any* $b \in R$ gives $f(ab) = f(a)f(b) = 0f(b) = 0$, and so multiplying any element in the kernel by an arbitrary element in the rings keeps it in the kernel. We would like to generalize these properties into an *ideal*.

Definition 5.12 (Ideals)

For a commutative ring $(R, +, \times)$, a **two-sided ideal**—or **ideal**—is a subset $I \subset R$ satisfying

1. $a, b \in I \implies a + b \in I$.
2. $a \in I, r \in R \implies ra = ar \in I$.[a]

If R is not necessarily commutative, then we $ra \neq ar$ in general, so we may distinguish between left and right ideals.

[a]Note that this property and closure under addition actually implies that it is an abelian subgroup under addition, since we can see that $-1 \in R$ and $a \in I$ implies $-1 \cdot a = -a \in I$.

Therefore, we can see that it is an abelian group under $+$ and closed under $\times$. However, it is not guaranteed to have a multiplicative identity, which is why we can interpret I as a ring without a multiplicative identity, also known as a *rung*. Ideals are analogous to normal subgroups, which were used to induce a congruence relation on a group to get its quotient. Ideals play a similar role.

Example 5.15 (Matrix with Last Row of Zeros)

Let R be the set of all $n \times n$ matrices. Then
1. The set of all $n \times n$ matrices whose last row is zero forms a right ideal, but not a left ideal.
2. The set of all $n \times n$ matrices whose last column is zero is a left ideal, but not a right ideal.

Example 5.16 (Multiples of Elements Are an Ideal)

We give 2 ideals:
1. The set of even integers $2\mathbb{Z}$ is an ideal in the ring $\mathbb{Z}$, since the sum of any even integers is even and the product of any even integer with an integer is an even integer. However, the odd integers do not form an ideal.
2. The set of all polynomials with real coefficients which are divisible by the polynomial $x^2 + 1$ is an ideal in the ring of all polynomials.

Let's talk about a few more properties of ideals, namely their construction and behavior under set theoretic operations.

Theorem 5.11 (Sum and Intersection of Ideals are Ideals)

Given two ideals $I, J \subset R$,
1. $I \cap J$ is an ideal.
2. $I + J \coloneqq \{i + j \mid i \in I, j \in J\}$ is an ideal.

Proof. Listed.
1. $I \cap J$ is an ideal. Given $a, b \in I \cap J$, then $a, b \in I \implies a + b \in I$, and $a, b \in J \implies a + b \in J$. So $a + b \in I \cap J$. Furthermore, for every $r \in R$, $a \in I \implies ra \in I$ and $a \in J \implies ra \in J$, so $a \in I \cap J \implies ra \in I \cap J$.
2. $I + J$ is an ideal. Given $x, y \in I + J$, then $x = a_x + b_x$ and $y = a_y + b_y$ for $a_x, a_y \in I, b_x, b_y \in J$. So
$$x + y = (a_x + b_x) + (a_y + b_y) = (a_x + a_y) + (b_x + b_y) \tag{136}$$
where $a_x + a_y \in I, b_x + b_y \in J$ by definition of an ideal, and so $x + y \in I + J$. Noe let $x = a_x + b_x \in I + J$. Then given $r \in R$,
$$rx = r(a_x + b_x) = ra_x + rb_x \tag{137}$$
where $ra_x \in I$ and $rb_x \in J$ since I, J are ideals. Therefore $rx \in I + J$.

Theorem 5.12 (Preimage of Ideals are Ideals)

If $f : R \to S$ is a ring homomorphism of commutative rings $J \subset S$ is an ideal, then $f^{-1}(J)$ is an ideal of R.

Proof. We prove the two properties of an ideal.
1. Consider $a, b \in f^{-1}(J) \subset R$. Then $f(a+b) = f(a) + f(b) \in J \implies a + b \in f^{-1}(J)$.
2. Consider $r \in R$ and $a \in If^{-1}(J)$. Then, $f(ra) = f(r)f(a)$ where $f(r) \in S$ and $f(a) \in J$. So $f(r)f(a) = f(ra) \in J \implies ra \in f^{-1}(J)$.

Example 5.17 (Image of Ideal is Not Necessarily an Ideal)

It is not true in general that for an ideal $I \subset R$ and a ring homomorphism $f : R \to S$, the image $f(I)$ is an ideal of S.

Given the two examples above, let's formalize the idea of an ideal consisting of all multiples of a specific element a. This sounds pretty familiar to *generators* of groups.

Definition 5.13 (Generators of Ideals)

Given a commutative ring R, the **ideal generated by** $a \in R$ is denoted

$$\langle a \rangle \coloneqq \{ra \mid r \in R\} \tag{138}$$

and more generally, we may have multiple generating elements.

$$\langle a_1, \dots, a_n \rangle \coloneqq \{r_1 a_1 + \dots r_n a_n \mid r_1, \dots, r_n \in R\} \tag{139}$$

A good—yet not completely accurate—intuition to have about ideals is that they are the set of multiples of a certain element. This technically isn't true in general, but if this intuition is true, then we call this a *principal ideal*.

Definition 5.14 (Principal Ideals)

A **principal ideal** is an ideal generated by a single element: $I = \langle a \rangle$.

Example 5.18 (Some Principal Ideals)

Let's take a look at some examples and non-examples of principal ideals.
1. In any ring R, the sets $\{0\} = \langle 0 \rangle$ and R are principal ideals.
2. The set of all even integers $2\mathbb{Z} = \{\dots, -4, -2, 0, 2, 4, \dots\}$ is a principal ideal generated by $2 \in \mathbb{Z}$.
3. The set of all imaginary multiples $i\mathbb{Z} = \{ai \mid a \in \mathbb{Z}\}$ is a principal ideal $\langle i \rangle \subset \mathbb{Z}[i]$.

What is nice about principal ideals it that they "encode" the divisors.

Theorem 5.13 (Principal Ideals and Divisors)

Given a commutative ring R and $a, b \in R$, the following are equivalent.
1. $b \mid a$.
2. $a \in \langle b \rangle$.

3. $\langle a \rangle \subset \langle b \rangle$.

Proof.

This allows us to define the GCD with ideals.

Corollary 5.14 (GCD as a Minimal Ideal)

Let R be a commutative ring, $a, b \in R$, and $I = \langle a, b \rangle$. Then d is the greatest common divisor if
1. I is contained in the principal ideal $\langle d \rangle$.
2. If $\langle d' \rangle$ is any principal ideal containing I, then $\langle d \rangle \subseteq \langle d' \rangle$.

Proof.

That is, a greatest common divisor of a, b is a generator for the smallest principal ideal containing a and b. However, there are cases in which the gcd is not unique, and hence there are multiple elements d, d' that generate such a minimal ideal! This can happen in two ways: either $\langle d \rangle = \langle d' \rangle$, or both principal ideals $\langle d \rangle, \langle d' \rangle$ contain a and b but are not contained in each other.

Example 5.19 (Two Distinct Elements can Generate the Same Ideal)

Continuing on our example from before, let's verify that the associate elements indeed generate the same ideal.
1. For $\pm 6 \in \mathbb{Z}$, we indeed have
$$\langle 6 \rangle = \{\dots, -12, -6, 0, 6, 12, \dots\} \tag{140}$$
$$\langle -6 \rangle = \{\dots, 12, 6, 0, -6, -12, \dots\} \tag{141}$$
2. For $1+i, 1-i \in \mathbb{Z}[i]$, every element in $\langle 1+i \rangle$ is of the form $r(1+i)$ for some $r \in \mathbb{Z}[i]$. Therefore,
$$r(1+i) = ri(-i)(1+i) = ri(1-i) \in \langle 1-i \rangle \tag{142}$$
for some $ri \in \mathbb{Z}[i]$, implying that $\langle 1+i \rangle \subset \langle 1-i \rangle$. The reverse inclusion is the same logic.
3. Given $2, 10 \in \mathbb{Z}_{12}$, let us write out their ideals.
$$\langle 2 \rangle = \{0, 2, 4, 6, 8, 10\} \tag{143}$$
$$\langle 10 \rangle = \{0, 10, 20 = 8, 18 = 6, 16 = 4, 14 = 2\} \tag{144}$$

This gives us a hint as to what these elements are.

Theorem 5.15 (Elements that Generate Same Ideal are Associated)

Given commutative ring R and $a, b \in R$, a and b are associate elements if and only if $\langle a \rangle = \langle b \rangle$.

Proof. We prove bidirectionally.
1. ($\rightarrow$). Let $a, b \in R$ be associate elements. Then there exists a unit $u \in R$ s.t. $a = ub$. Therefore $\langle a \rangle \subset \langle b \rangle$. On the other hand, $b = u^{-1}a$, and so $\langle b \rangle \subset \langle a \rangle$.
2. ($\leftarrow$). Let $\langle a \rangle = \langle b \rangle$. Then this implies that $a \mid b$ and $b \mid a$, and so they are associate elements.

Now that we have learned ideals, we can define prime ideals and prime elements.

Definition 5.15 (Prime Ideal)

An ideal P of a commutative ring R is **prime** if it has the following two properties.
1. If $a, b \in R$ where $ab \in P$, then either $a \in P$ or $b \in P$.
2. $P \subsetneq R$.

Definition 5.16 (Prime Elements)

Let R be a commutative ring and $p \in R$ be nonzero and not a unit. p is **prime** if either of the equivalent conditions is true.
1. The ideal $\langle p \rangle$ is a prime ideal.
2. Whenever $p \mid ab$ for any $a, b \in R$, then either $p \mid a$ or $p \mid b$.

The final property of ideals is whether it "almost" fills up the whole ring in that it is maximal.

Definition 5.17 (Maximal Ideal)

Let R be a commutative ring and $I \subset R$ an ideal. I is **maximal** if there exists no ideal J s.t. $I \subsetneq J \subsetneq R$.

5.5 Quotient Rings

What is nice about ideals is that they induce not just an equivalence relation—but a congruence relation—on a ring, which is a generalization of working in the integers modulo n.

Theorem 5.16 (Equivalence Relation Induced by an Ideal)

Given a commutative ring R and an ideal $I \subset R$, we say that two elements $a, b \in R$ are **congruent** (mod I), written $a \equiv b \pmod{I}$ iff $a - b \in I$. We claim two things:
1. $\equiv$ is an equivalence relation.
2. $\equiv$ is a congruence relation. Given that $a \equiv a' \pmod{I}$ and $b \equiv b' \pmod{I}$,

$$a + b \equiv a' + b' \pmod{I}, \qquad ab \equiv a'b' \pmod{I} \tag{145}$$

Occasionally, if the ideal I is clear from context, we will write $a \equiv b$.

Proof. We first prove that $\equiv$ is indeed an equivalence relation.
1. *Reflexive.* $a \equiv a \pmod{I}$ is trivial since $a - a = 0 \in I$.
2. *Symmetric.* If $a \equiv b$, then $a - b \in I \implies -(a - b) = -a + b = b - a \in I \implies b \equiv a$.
3. *Transitive.* If $a \equiv b$ and $b \equiv c$, then $a - b \in I$ and $b - c \in I$. Since I is an additive group and so it is closed under addition, so $(a - b) + (b - c) = a - c \in I \implies a \equiv c$.

Note that so far, we have only used the group property of ideals to prove that is is an equivalence class. Now for congruence of multiplication, we need the ring properties.
1. $a \equiv a', b \equiv b' \implies (a - a'), (b - b') \in I$. By adding them together and distributivity, we have

$$a - a' + b - b' = (a + b) - (a' + b') \in I \implies a + b \equiv a' + b' \pmod{I} \tag{146}$$

2. We see that $a \in R, (b - b') \in I \implies a(b - b') \in I$. Similarly, $b' \in R, (a - a') \in I \implies (a - a')b' \in I$. Now adding the two, we have

$$a(b - b') + (a - a')b' = ab - ab' + ab' - a'b' = ab - a'b' \in I \implies ab = a'b' \pmod{I} \tag{147}$$

This quotient space maintains a lot of nice properties of the algebraic operations, and so we can form a new

ring structure with this quotient space.

Definition 5.18 (Quotient Rings, Rings of Residue Class)

The quotient space R/I induced by the mapping $a \mapsto [a]$ is indeed a commutative ring, called the **quotient ring**, with addition and multiplication defined

$$[a] + [b] \coloneqq [a+b], \qquad [ab] \coloneqq [a]\,[b] \tag{148}$$

Proof. Note that the properties of the operation in $\frac{M}{R}$ inherits all the properties of the addition operation on M that are expressed in the form of identities and inverses, along with the existence of the zero identity.

$$\begin{aligned} 0 \in M &\implies [0] \text{ is the additive identity in } \frac{M}{R} \\ a + (-a) = 0 &\implies [a] + [-a] = [0] \\ 1 \in M &\implies [1] \text{ is the multiplicative identity in } \frac{M}{R} \end{aligned}$$

Theorem 5.17 (Quotient Maps are Homomorphisms)

The map $p : R \to R/I$ is a ring homomorphism.

Proof. This is true by definition since we have made $\equiv$ a congruence relation.

Example 5.20 (Quotient Rings of Integers)

The quotient set $\mathbb{Z}/\langle n \rangle$ by the relation of congruence modulo n is denoted $\mathbb{Z}_n$.

$$\mathbb{Z}_n = \{[0]_n, [1]_n, \dots, [n-1]_n\} \tag{149}$$

Note that the quotient ring $(\mathbb{Z}/\langle n \rangle, +, \times)$ is precisely the cyclic quotient group $\mathbb{Z}_n = \mathbb{Z}/6\mathbb{Z}$ when considering only addition. We list some quotient rings of the integers.

1. In $\mathbb{Z}_5 = \mathbb{Z}/\langle 5 \rangle$, the elements [2] and [3] are multiplicative inverses of each other since $[2][3] = [6] = [1]$, and [4] is its own inverse since $[4][4] = [16] = [1]$. The addition and multiplication tables for $\mathbb{Z}_5$ is shown below.
2. Consider the ideal $I = \langle 2 \rangle \subset \mathbb{Z}_6$. We have $0 \equiv 2 \equiv 4 \pmod{I}$ and $1 \equiv 3 \equiv 5 \pmod{I}$, and so the quotient ring $\mathbb{Z}_6/I$ consists of the two equivalence classes [0] and [1].

Example 5.21 (Quotient Rings of Polynomials)

We list some quotient rings of polynomials.

1. Consider $\mathbb{Q}[x]/\langle x^2 - 2 \rangle$. We can see that any polynomial $f \in \mathbb{Q}[x]$ is equivalent $\pmod{I}$ to a linear polynomial, since $x^2 \equiv 2$. Alternatively we can apply the division algorithm to replace $f(x)$ by its remainder upon division by $x^2 - 2$, and thus in the quotient ring, $[x]$ plays the role of $\sqrt{2}$, which may indicate that $\mathbb{Q}[x]/\langle x^2 - 2 \rangle = \mathbb{Q}[\sqrt{2}]$.
2. Consider $\mathbb{Z}_2[x]/\langle x^2 + x + 1 \rangle$. As in the previous example, any polynomial in $\mathbb{Z}_2[x]$ is equivalent to a linear polynomial since $x^2 \equiv x + 1 \pmod{I}$. Therefore the elements of the quotient ring are $[0], [1], [x], [x+1]$ with the addition and multiplication tables.

$+$	0	1	x	$x+1$
0	0	1	x	$x+1$
1	1	0	$x+1$	x
x	x	$x+1$	0	1
$x+1$	$x+1$	x	1	0

(a)

$\cdot$	0	1	x	$x+1$
0	0	0	0	0
1	0	1	x	$x+1$
x	0	x	$x+1$	1
$x+1$	0	$x+1$	1	x

(b)

Note that just like how quotient topologies do not preserve topological properties, as shown here and here, quotient rings inherit some—but not all—ring properties. It obviously inherits commutativity since the quotient space is constructed with a congruence relation. However, the characteristic is changed.

Just like in group theory, we have a method of constructing isomorphisms between cleverly chosen rings S and a quotient ring R/I. This seems to be a common pattern here when considering groups, rings, and topological spaces... This will be investigated more in category theory.

Theorem 5.18 (First Isomorphism Theorem for Rings)

Let R and S be commutative rings, and suppose $f : R \to S$ be a surjective ring homomorphism. Then this induces a ring isomorphism

$$R/\ker f \simeq S \tag{150}$$

satisfying $\phi = \bar{\phi} \circ \pi$.

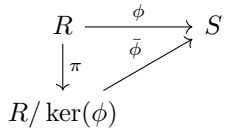


Figure 11: The theorem states that the following diagram commutes.

Proof.

A direct application of this is the Chinese remainder theorem.

Corollary 5.19 (Chinese Remainder Theorem)

Given a commutative ring R, let $I, J \subset R$ be ideals such that $I + J = R$. Then,

$$\pi : R \to \frac{R}{I} \times \frac{R}{J}, \qquad r \mapsto ([r]_I, [r]_J) \tag{151}$$

with component-wise quotient mappings is a surjective ring homomorphism with $\ker \pi = I \cap J$. By the fundamental ring homomorphism theorem, it immediately follows that

$$\frac{R}{I \cap J} \simeq \frac{R}{I} \times \frac{R}{J} \tag{152}$$

Proof. Since $I + J = R$, there exists $i \in I$ and $j \in j$ s.t. $i + j = 1$. Let $\bar{a} = a + I \in R/I$ and $\bar{b} = b + J \in R/J$ be any elements. Then

$$\pi(aj + bi) = ([aj + bi]_I, [a_j + bi]_J) = ([aj]_I, [bi]_J) \in \frac{R}{I} \times \frac{R}{J} \tag{153}$$

But we have

1. $a(j+i) = a \in R \implies aj = a(j+i) \in R/I$. Therefore $[aj]_I = [a]_I$
2. $b(j+i) = b \in R \implies bi = bj + bi \in R/J$. Therefore $[b]_J = [bi]_J$.

Therefore, we have $\pi(aj + bi) = ([a]_I, [b]_J)$, which proves surjectivity.

Example 5.22

We claim that $\mathbb{Z}_{10} \simeq \mathbb{Z}_5 \times \mathbb{Z}_2$ as rings. In fact, the whole isomorphism is defined with the mappings $f(1,1) = 1$.

Example 5.23 (Chinese Remainder Theorem on Integers)

Suppose that we are solving the system of linear conguence equations

$$x \equiv b_1 \pmod{m_1}, \qquad x \equiv b_2 \pmod{m_2} \tag{154}$$

where m_1, m_2 are coprime. The Chinese remainder theorem says that there exists a solution x where any two solutions x_1, x_2 are congruent modulo N. We can think of each equation as modeling x as living in an ideal specified by the isomorphism

$$\phi : \frac{\mathbb{Z}}{N\mathbb{Z}} \to \frac{\mathbb{Z}}{m_1\mathbb{Z}} \times \frac{\mathbb{Z}}{m_2\mathbb{Z}}, \qquad x \pmod{N} \xrightarrow{\phi} (x \pmod{m_1}, x \pmod{m_2}) \tag{155}$$

Therefore, for doing a sequence of arithmetic operations in $\mathbb{Z}/N\mathbb{Z}$, we can do in each component ring and then get the result by applying the isomorphism backwards. This isomorphism becomes

$$x \equiv a_2 b_1 m_2 + a_1 b_2 m_1 \pmod{m_1 m_2} \tag{156}$$

where $1 = a_1 m_1 + a_2 m_2$. Alternatively, we can start off with the congruences, and begin by using Bezout's identity on m_1, m_2, and then multiplying by the b_1, b_2.

$$1 = a_1 n + a_2 m_2 \tag{157}$$
$$b_1 = a_1 b_1 m_1 + a_2 b_1 m_2 \tag{158}$$
$$b_2 = a_1 b_2 m_1 + a_2 b_2 m_2 \tag{159}$$

Then we can cleverly set

$$x = a_2 b_1 m_2 + a_1 b_2 m_1 \tag{160}$$

which now satisfies $x \equiv a_2 b_1 m_2 \pmod{m_1}$ and $x \equiv a_1 b_2 m_1 \pmod{m_2}$.

5.6 Division Rings

Definition 5.19 (Division Ring)

A **division ring**, also called a **skew field**, is an associative ring where every nonzero element is invertible with respect to $\times$.[a]

[a] Division rings differ from fields in that multiplication is not required to be commutative.

Let's establish the hierarchy.

Lemma 5.20 (Division Rings are Domains)

Every division ring R is automatically a domain.

Proof. Every nonzero element is a unit and hence cannot be a zero-divisor.

Example 5.24 (Invertible Matrices)

The classic example is the ring of invertible matrices over the reals $\mathrm{GL}(\mathbb{R}^n)$, which is not necessarily commutative, but is a ring in which "division" can be done by right and left multiplication of a matrix inverse.

$$AA^{-1} = A^{-1}A = I \tag{161}$$

This implies that every element in the division ring commutes with the identity, but again commutativity does not necessarily hold for arbitrary elements A, B.

Example 5.25 (Hamiltonian Quaternions)

The real Hamiltonian quaternions

$$\mathbb{H} \coloneqq \{a + bi + cj + dk \mid a, b, c, d \in \mathbb{R}\} \tag{162}$$

where addition is defined component-wise and multiplication defined by expanding

$$i^2 = j^2 = k^2 = 1, \quad ij = -ji = k, \quad jk = -kj = k, \quad ki = -ik = j \tag{163}$$

where the real coefficients commute with i, j, k. This can be proven tediously to be a division ring, along with the rational Hamiltonian quaternions.

5.7 Exercises

Exercise 5.1 (Shifrin 1.2.1)

For each of the following pairs of numbers a and b, find $d = \gcd(a, b)$ and express d in the form $ma + nb$ for suitable integers m and n.

(a) $14, 35$
(b) $56, 77$
(c) $618, 336$
(d) $2873, 6643$
(e) $512, 360$
(f) $4432, 1080$

Solution. Listed.

1. $d = 7 = (-2) \cdot 14 + (1) \cdot 35$.
2. $d = 7 = (-4) \cdot 56 + 3 \cdot 77$.
3. $d = 6 = -25 \cdot 618 + 46 \cdot 336$
4. $d = 13 = 37 \cdot 2873 + (-16) \cdot 6643$.
5. $d = 8 = 19 \cdot 512 + (-27) \cdot 360$.
6. $d = 8 = 29 \cdot 4432 + (-119) \cdot 1080$.

Exercise 5.2 (Shifrin 1.2.2)

You have at your disposal arbitrarily many 4-cent stamps and 7-cent stamps. What are the postages you can pay? Show in particular that you can pay all postages greater than 17 cents.

Solution.

Exercise 5.3 (Shifrin 1.2.3)

Prove that whenever $m \neq 0$, $\gcd(0, m) = |m|$.

Solution.

Exercise 5.4 (Shifrin 1.2.4)

(a) Prove that if $a|x$ and $b|y$, then $ab|xy$.
(b) Prove that if $d = \gcd(a, b)$, then $\gcd(\frac{a}{d}, \frac{b}{d}) = 1$.

Solution.

Exercise 5.5 (Shifrin 1.2.5)

Prove or give a counterexample: the integers q and r guaranteed by the division algorithm, Theorem 2.2, are unique.

Solution.

Exercise 5.6 (Shifrin 1.2.6)

Prove or give a counterexample. Let $a, b \in \mathbb{Z}$. If there are integers m and n so that $d = am + bn$, then $d = \gcd(a, b)$.

Solution.

Exercise 5.7 (Shifrin 1.2.7)

Generalize Proposition 2.5: if $\gcd(m, c) = 1$ and $m|cz$, then prove $m|z$.

Solution. Let $\gcd(m, c) = 1$ and $m|cz$. Then there exists $a, b \in \mathbb{Z}$ such that $am + bc = 1$. Multiply both sides of the equation by z to get by the distributive property

$$(am + bc)z = amz + bcz = z \tag{164}$$

$m|amz$ and $m|cz \implies m|bcz$. Therefore, the sum of the two, which is equal to z, must be divisible by m. Therefore $m|z$.

Exercise 5.8 (Shifrin 1.2.8)

Suppose $a, b, n \in \mathbb{N}$, $\gcd(a, n) = 1$, and $\gcd(b, n) = 1$. Prove or give a counterexample: $\gcd(ab, n) = 1$.

Solution.

Exercise 5.9 (Shifrin 1.2.9)

Prove that if p is prime and $p|(a_1 a_2 \dots a_n)$, then $p|a_j$ for some j, $1 \leq j \leq n$. (Hint: Use Proposition 2.5 and induction.)

Solution.

Exercise 5.10 (Shifrin 1.2.10)

Given a positive integer n, find n consecutive composite numbers.

Solution.

Exercise 5.11 (Shifrin 1.2.11)

Prove that there are no integers m, n so that $(\frac{m}{n})^2 = 2$. (Hint: You may start by assuming m and n are relatively prime. Why? Then use Exercise 1.1.3.)

Solution.

Exercise 5.12 (Shifrin 1.2.12)

Find all rectangles whose sides have integral lengths and whose area and perimeter are equal.

Solution.

Exercise 5.13 (Shifrin 1.2.13)

Given two nonzero integers a, b, in analogy with the definition of $\gcd(a, b)$, we define the **least common multiple** $\text{lcm}(a, b)$ to be the positive number μ with the properties:

(i) $a|\mu$ and $b|\mu$, and
(ii) if $s \in \mathbb{Z}$, $a|s$ and $b|s \Rightarrow \mu|s$.

Prove that

(a) if $\gcd(a, b) = 1$, then $\mu = ab$. (Hint: If $\gcd(a, b) = 1$, then there are integers m and n so that $1 = ma + nb$; therefore, $s = mas + nbs$.)
(b) more generally, if $\gcd(a, b) = d$, then $\mu = ab/d$.

Solution. Listed.

1. We can simply verify the two properties. Since $\mu = ab$, $a|\mu$ and $b|\mu$ trivially by the existence of b and a, respectively. As for the second property, let $s \in \mathbb{Z}$ exist such that $a|s$ and $b|s$. Since $a|s$, $s = xa$ for some $x \in \mathbb{Z}$. But since $b|s$, $b|xa$. Since $\gcd(a, b) = 1$ by assumption, the result in [Shifrin 1.2.7] tells us that $b|x$, i.e. there exists some $k \in \mathbb{Z}$ such that $x = kb$. Therefore

$s = xa = kba = kab = k\mu$. By existence of k, $\mu|s$, and we are done.

2. Given a, b with $\gcd(a,b) = d$, there exists some $a', b' \in \mathbb{Z}$ s.t. $a = da', b = db'$. We claim that $\mu = ab/d \coloneqq da'b'$ is the lcm.[a] It is clear that $a|\mu$ and $b|\mu$ by the existence of integers b' and a', respectively. To prove the second property, let $s \in \mathbb{Z}$ with $a|s$ and $b|s$. Since $a|s \iff da'|s$, there must exist some $x \in \mathbb{Z}$ s.t. $s = da'x$. But since $b|s$, this means that $db'|s \iff db'|da'x \iff b'|a'x$. But $\gcd(a', b') = 1$ which follows from the definition of gcd, and so by [Shifrin 1.2.7] it must be the case that $b'|x$, i.e. there exists some $k \in \mathbb{Z}$ s.t. $x = b'k$. Substituting this back we have $s = da'b'k = \mu k$, and by existence of k it follows that $\mu|s$. Since it satisfies these 2 properties μ is the lcm.

[a] Since division isn't generally closed in the integers, I prefer to define ab/d this way.

Exercise 5.14 (Shifrin 1.2.14)

See Exercise 13 for the definition of $\text{lcm}(a,b)$. Given prime factorizations $a = p_1^{\mu_1} \cdots p_m^{\mu_m}$ and $b = p_1^{\nu_1} \cdots p_m^{\nu_m}$, with $\mu_i, \nu_i \geq 0$, express $\gcd(a,b)$ and $\text{lcm}(a,b)$ in terms of $p_1, \dots, p_m$. Prove that your answers are correct.

Solution.

Exercise 5.15 (Shifrin 1.3.8)

Determine the last digit of 3^{400}; then the last two digits. Determine the last digit of 7^{99}.

Solution. We see that in mod10,

$$3^{400} \equiv 9^{200} \equiv (-1)^{200} \equiv 1^{100} \equiv 1 \tag{165}$$

so the last digit is 1. To get the last 2 digits, we use the binomial expansion and focus on the last 2 terms.

$$3^{400} = 9^{200} = (10-1)^{200} = \dots + \binom{200}{199} 10^1 (-1)^{199} + \binom{200}{200} (-1)^{200} \tag{166}$$

since every combination of the form $\binom{n}{k}$ is an integer and all the other terms have a factor of 10^2, the expansion mod100 becomes

$$3^{400} \equiv \binom{200}{199} 10^1 (-1)^{199} + \binom{200}{200} (-1)^{200} = 200 \cdot 10 \cdot (-1)^{199} + 1 \equiv 1 \pmod{100} \tag{167}$$

and so the last two digits is 01. To get the last digit of 7^{99}, we see that in mod10,

$$7^{99} \equiv 7^{96} \cdot 7^3 \equiv (7^4)^{24} \cdot 343 \equiv 2401^{24} \cdot 343 \equiv 1^{24} \cdot 3 \equiv 3 \tag{168}$$

Exercise 5.16 (Shifrin 1.3.9)

Show that if an integer is a sum of two fourth powers, then its units digit must be 0, 1, 2, 5, 6, or 7.

Solution.

Exercise 5.17 (Shifrin 1.3.10)

Let $n = \sum_{i=0}^{k} a_i 10^i$. Prove that $13|n \iff 13|\left(\sum_{i=1}^{k} a_i 10^{i-1} + 4a_0\right)$. (Hint: Cf. the divisibility test for 7 in Proposition 3.2.)

Solution. We must show that

$$n \equiv 0 \pmod{13} \iff n' = \sum_{i=1}^{k} a_i 10^{i-1} + 4a_0 \equiv 0 \pmod{13} \tag{169}$$

We see that $n \equiv n + 39a_0 \equiv 0 \pmod{13}$, and

$$\begin{aligned} n + 39a_0 &= \sum_{i=0}^{k} 10^i a_i + 39a_0 && (170)\\ &= \sum_{i=1}^{k} 10^i a_i + 40a_0 && (171)\\ &= 10\left(\sum_{i=1}^{k} 10^{i-1} a_i + 4a_0\right) && (172)\\ &= 10n' && (173) \end{aligned}$$

and so we have $n \equiv 10n' \pmod{13}$, and so $n' \equiv 0 \pmod{13} \implies n \equiv 0 \pmod{13}$. Conversely, if $n \equiv 0 \pmod{13}$, then $4n \equiv 0 \pmod{13}$, but $4n \equiv 40n'$ and so $n' \equiv 40n' \equiv 4n \equiv 0 \pmod{13}$. Therefore both implications are proven.

Exercise 5.18 (Shifrin 1.3.12)

Suppose that p is prime. Prove that if $a^2 \equiv b^2 \pmod{p}$, then $a \equiv b \pmod{p}$ or $a \equiv -b \pmod{p}$.

Solution. We have

$$\begin{aligned} a^2 \equiv b^2 \pmod{p} &\implies a^2 - b^2 \equiv 0 \pmod{p} && (174)\\ &\implies (a+b)(a-b) \equiv 0 \pmod{p} && (175) \end{aligned}$$

We claim that there are no zero divisors in $\mathbb{Z}_p$. If $mn \equiv 0 \pmod{p}$, then by definition this means $p|mn$, which implies that in the integers this must mean that $p|m$ or $p|n$.[a] But since $m, n \not\equiv 0$, $p \not| n$ and $p \not| m$, arriving at a contradiction. Going back to our main argument, it must be the case that $a + b \equiv 0 \implies a \equiv -b$ or $a - b \equiv 0 \implies a \equiv b$.

[a] Proposition 2.5

Exercise 5.19 (Shifrin 1.3.15)

Show that if $n \equiv 7 \pmod{8}$, then n is not the sum of three squares.Show that if $n \equiv 7 \pmod{8}$, then n is not the sum of three squares.

Solution. Let us assume that $n = a^2 + b^2 + c^2$ for some $a, b, c \in \mathbb{Z}$. Let us consider for each integer z,

all the possible values of $z^2 \pmod 8$.

$$z \equiv 0 \implies z^2 \equiv 0 \pmod 8 \tag{176}$$
$$z \equiv 1 \implies z^2 \equiv 1 \pmod 8 \tag{177}$$
$$z \equiv 2 \implies z^2 \equiv 4 \pmod 8 \tag{178}$$
$$z \equiv 3 \implies z^2 \equiv 1 \pmod 8 \tag{179}$$
$$z \equiv 4 \implies z^2 \equiv 0 \pmod 8 \tag{180}$$
$$z \equiv 5 \implies z^2 \equiv 1 \pmod 8 \tag{181}$$
$$z \equiv 6 \implies z^2 \equiv 4 \pmod 8 \tag{182}$$
$$z \equiv 7 \implies z^2 \equiv 1 \pmod 8 \tag{183}$$

Therefore, $a^2 + b^2 + c^2 \pmod 8$ can take any values of the form

$$x + y + z \pmod 8 \text{ for } x, y, z \in \{0, 1, 4\} \tag{184}$$

Since addition is commutative, WLOG let $x \leq y \leq z$. We can just brute force search this.

1. If $z = 0$, then $x = y = z = 0$ and $x + y + z = 0 \not\equiv 7$.
2. If $z = 1$, then we see

$$0 + 0 + 1 \equiv 1 \tag{185}$$
$$0 + 1 + 1 \equiv 2 \tag{186}$$
$$1 + 0 + 1 \equiv 2 \tag{187}$$
$$1 + 1 + 1 \equiv 3 \tag{188}$$

3. If $z = 4$, then we see that

$$0 + 0 + 4 \equiv 4 \tag{189}$$
$$0 + 1 + 4 \equiv 5 \tag{190}$$
$$0 + 4 + 4 \equiv 0 \tag{191}$$
$$1 + 1 + 4 \equiv 6 \tag{192}$$
$$1 + 4 + 4 \equiv 1 \tag{193}$$
$$4 + 4 + 4 \equiv 4 \tag{194}$$

And so $a^2 + b^2 + c^2 \not\equiv 7 \pmod 8$ for any $a, b, c \in \mathbb{Z}$.

Exercise 5.20 (Shifrin 1.3.20)

Solve the following congruences:

a. $3x \equiv 2 \pmod 5$
b. $6x + 3 \equiv 1 \pmod{10}$
c. $243x + 17 \equiv 101 \pmod{725}$
d. $20x \equiv 4 \pmod{30}$
e. $20x \equiv 30 \pmod 4$
f. $49x \equiv 4000 \pmod{999}$
g. $15x \equiv 25 \pmod{35}$
h. $15x \equiv 24 \pmod{35}$

Solution. For (a),

$$3x \equiv 2 \pmod 5 \implies 6x \equiv 4 \pmod 5 \implies x \equiv 4 \pmod 5 \tag{195}$$

For (b),

$$\begin{aligned} 6x + 3 \equiv 1 \pmod{10} &\implies 6x \equiv -2 \equiv 8 \pmod{10} && (196)\\ &\implies 10|(6x-8) && (197)\\ &\implies 5|(3x-4) && (198)\\ &\implies 3x \equiv 4 \pmod 5 && (199)\\ &\implies 3x \equiv 9 \pmod 5 && (200)\\ &\implies x \equiv 3 \pmod 5 && (201) \end{aligned}$$

For (g),

$$\begin{aligned} 15x \equiv 25 \pmod{35} &\implies 35|(15x-25) && (202)\\ &\implies 7|(3x-5) && (203)\\ &\implies 3x \equiv 5 \pmod 7 && (204)\\ &\implies 3x \equiv 12 \pmod 7 && (205)\\ &\implies x \equiv 4 \pmod 7 && (206) \end{aligned}$$

Exercise 5.21 (Shifrin 1.3.21)

Use the Chinese Remainder Theorem, Theorem 3.7 (or its improvement, Theorem 3.8), to solve the following simultaneous congruences:

a. $x \equiv 1 \pmod 3$, $x \equiv 1 \pmod 5$
b. $x \equiv 1 \pmod 4$, $x \equiv 7 \pmod{13}$
c. $x \equiv 3 \pmod 4$, $x \equiv 4 \pmod 5$, $x \equiv 3 \pmod 7$
d. $19x \equiv 1 \pmod{140}$ (Hint: factor 140 and see Exercise 11.)
e. $x \equiv 3 \pmod 9$, $x \equiv 18 \pmod{24}$
f. $x \equiv 4 \pmod{105}$, $x \equiv 29 \pmod{80}$
g. $x \equiv 11 \pmod{142}$, $x \equiv 25 \pmod{86}$

Solution. For (b), we see that 4 and 13 are coprime with $-3 \cdot 4 + 1 \cdot 13 = 1$. Therefore, by the Chinese remainder theorem

$$x \equiv 1 \cdot 1 \cdot 12 + (-3) \cdot 7 \cdot 4 \pmod{52} \implies x \equiv 33 \pmod{52} \tag{207}$$

For (c), we solve the first two congruences $x \equiv 3 \pmod 4$ and $x \equiv 4 \pmod 5$. 4 and 5 are coprime with $-1 \cdot 4 + 1 \cdot 5 = 1$. Therefore, by CRT

$$x \equiv -1 \cdot 4 \cdot 4 + 1 \cdot 5 \cdot 3 \pmod{20} \implies x \equiv -1 \pmod{20} \tag{208}$$

Then we solve $x \equiv -1 \pmod{20}$ with the final congruence $x \equiv 3 \pmod 7$. We see that 20 and 7 are coprime with $-1 \cdot 20 + 3 \cdot 7 = 1$. Therefore by CRT

$$x \equiv -1 \cdot 20 \cdot 3 + 3 \cdot 7 \cdot -1 \pmod{140} \implies x \equiv 59 \pmod{140} \tag{209}$$

Exercise 5.22 (Shifrin 1.3.25)

Prove that one can solve the congruence $cx \equiv b \pmod{m}$ if and only if $\gcd(c, m)|b$. Show, moreover, that the answer is unique modulo $m/\gcd(c, m)$.

Solution. We prove bidirectionally.

1. Assume a solution exists for $cx \equiv b \pmod{m}$. Then $m|(cx - b)$, which means that there exists a $y \in \mathbb{Z}$ s.t. $my = cx - b \iff b = cx - my$. Since $d = \gcd(c, m)$, there exists $c', m' \in \mathbb{Z}$ s.t. $c = dc'$ and $m = dm'$. So
$$b = cx - my = d(c'x - m'y) \implies d|b \tag{210}$$
2. Assume that $d|b$. Then there exists a $b' \in \mathbb{Z}$ s.t. $b = db'$, and we have
$$\begin{aligned} cx \equiv b \pmod{m} &\iff m|(cx - b) && (211)\\ &\iff dm'|d(c'x - b') && (212)\\ &\iff m'|(c'x - b') && (213)\\ &\iff c'x \equiv b' \pmod{m'} && (214)\end{aligned}$$
Since $\gcd(c', m') = 1$[a], by Shifrin Proposition 3.5 the equation $c'x \equiv b' \pmod{m'}$ is guaranteed to have a solution, and working backwards in the iff statements gives us the solution for $cx \equiv b \pmod{m}$.

We have proved existence of a solution in $\mod(m/d) = m'$. Now we show uniqueness. Assume that there are two solutions $x \equiv \alpha$, $x \equiv \beta \pmod{m'}$ with $\alpha \not\equiv \beta \pmod{m'}$. Then, x can be written as $x = k_\alpha m' + \alpha$ and $x = k_\beta m' + \beta$. But we see that
$$\begin{aligned} 0 = x - x &= (k_\alpha m' + \alpha) - (k_\beta m' + \beta) && (215)\\ &= m'(k_\alpha - k_\beta) + (\alpha - \beta) && (216)\\ &\equiv \alpha - \beta \pmod{m'} && (217)\end{aligned}$$
which implies that $\alpha \equiv \beta \pmod{m'}$, contradicting our assumption that they are different in modulo. Therefore the solution must be unique.

[a] Since $\gcd(c, m) = d \implies$ that there exists a $y, z \in \mathbb{Z}$ s.t. $cy + mz = d$, and dividing both sides by d guarantees the existence of y, z satisfying $c'y + m'z = 1$, meaning that $\gcd(c', m') = 1$.

Exercise 5.23 (Shifrin 1.4.1)

Construct multiplication tables for $\mathbb{Z}_7$, $\mathbb{Z}_8$, $\mathbb{Z}_9$, and $\mathbb{Z}_{12}$. List the zero-divisors and units in each ring.

Solution. For $\mathbb{Z}_7$. There are no zero divisors and the units are all elements.

$\times$	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6
2	0	2	4	6	1	3	5
3	0	3	6	2	5	1	4
4	0	4	1	5	2	6	3
5	0	5	3	1	6	4	2
6	0	6	5	4	3	2	1

(218)

For $\mathbb{Z}_8$. The zero divisors are $2, 4, 6$. The units are $1, 3, 5, 7$.

$\times$	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7
2	0	2	4	6	0	2	4	6
3	0	3	6	1	4	7	2	5
4	0	4	0	4	0	4	0	4
5	0	5	2	7	4	1	6	3
6	0	6	4	2	0	6	4	2
7	0	7	6	5	4	3	2	1

(219)

For $\mathbb{Z}_{12}$. The zero divisors are $2, 3, 4, 6, 8, 9, 10$. The units are $1, 5, 7, 11$.

$\times$	0	1	2	3	4	5	6	7	8	9	10	11
0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7	8	9	10	11
2	0	2	4	6	8	10	0	2	4	6	8	10
3	0	3	6	9	0	3	6	9	0	3	6	9
4	0	4	8	0	4	8	0	4	8	0	4	8
5	0	5	10	3	8	1	6	11	4	9	2	7
6	0	6	0	6	0	6	0	6	0	6	0	6
7	0	7	2	9	4	11	6	1	8	3	10	5
8	0	8	4	0	8	4	0	8	4	0	8	4
9	0	9	6	3	0	9	6	3	0	9	6	3
10	0	10	8	6	4	2	0	10	8	6	4	2
11	0	11	10	9	8	7	6	5	4	3	2	1

(220)

Exercise 5.24 (Shifrin 1.4.5.a/b/c)

1. Prove that $\gcd(a, m) = 1 \iff \bar{a} \in \mathbb{Z}_m$ is a unit.
2. Prove that if $\bar{a} \in \mathbb{Z}_m$ is a zero-divisor, then $\gcd(a, m) > 1$, and conversely, provided $m \nmid a$.
3. Prove that every nonzero element of $\mathbb{Z}_m$ is either a unit or a zero-divisor.
4. Prove that in any commutative ring R, a zero-divisor cannot be a unit, and a unit cannot be a zero-divisor. Do you think c. holds in general?

Solution. For (a),

1. ($\rightarrow$). If $\gcd(a, m) = 1$, then there exists $x, y \in \mathbb{Z}$ such that $ax + my = 1$. Taking the modulo on both sides gives $ax \equiv 1 \pmod{m}$, and therefore we have established the existence of $x \in \mathbb{Z}$, which implies the existence of $\bar{x} \in \mathbb{Z}_m$.
2. ($\leftarrow$). If we have $a \in \mathbb{Z}$ and $\bar{a}$ is a unit, then there exists a $\bar{x} \in \mathbb{Z}_m$ s.t. $\bar{a}\bar{x} = \bar{1} \iff ax \equiv 1 \pmod{m}$, which means that $m|(1 - ax)$. So there exists an integer $y \in \mathbb{Z}$ s.t. $my = 1 - ax \iff ax + my = 1$. By Shifrin corollary 2.4 a, m must be coprime.

For (b),

1. ($\rightarrow$) Let $\bar{a} \in \mathbb{Z}_m$ be a zero-divisor. Then there exists $\bar{x} \neq \bar{0}$ in $\mathbb{Z}_m$ such that $\bar{a}\bar{x} = \bar{0}$. This means: $ax \equiv 0 \pmod{m}$, so $m \mid ax$, and $m \nmid x$ (since $\bar{x} \neq \bar{0}$). Since $m \mid ax$ but $m \nmid x$, some prime factor of m must divide a. This prime factor is then a common divisor of a and m greater than 1, so $\gcd(a, m) > 1$.
2. ($\leftarrow$) Let $a \in \mathbb{Z}$, $m \in \mathbb{N}$ where $\gcd(a, m) = d > 1$ and $m \nmid a$. Then $a = a'd$ and $m = m'd$ for some $a', m' \in \mathbb{Z}$. Therefore,
$$\bar{a}\bar{m'} = \overline{am'} = \overline{a'dm'} = \overline{a'm} = \bar{0} \tag{221}$$
Also since $m \nmid a$, we have $\bar{a} \neq \bar{0}$, and since $m = m'd$, we have $m \nmid m'$ (since $m \nmid a \implies d \neq m$),

so $\bar{m'} \neq \bar{0}$. Therefore $\bar{a}$ is a zero-divisor in $\mathbb{Z}_m$.

For (c), let $a \in \mathbb{Z}_m$ be a nonzero element. Then it must be the case that $\gcd(a, m) = 1$ or $\gcd(a, m) > 1$. In the former case, a is a unit by (a), and in the latter case, $a \not\equiv 0 \implies m \nmid a$[a], and so by (b) a is a zero divisor.

[a] By contrapositive $m \mid a \implies a \equiv 0 \pmod{m}$ is trivial.

Exercise 5.25 (Shifrin 1.4.6.b/c/d)

Prove that in any ring R:
1. $0 \cdot a = 0$ for all $a \in R$ (cf. Lemma 1.1);
2. $(-1)a = -a$ for all $a \in R$ (cf. Lemma 1.2);
3. $(-a)(-b) = ab$ for all $a, b \in R$;
4. the multiplicative identity $1 \in R$ is unique.

Solution. For (a), note that $0a = (0 + 0) \cdot a = 0a + 0a$ and by subtracting $0a$ from both sides, we have $0 = 0a$. Similarly, $a0 = a(0 + 0) = a0 + a0 \implies 0 = a0$. For (b),

$$\begin{aligned} a + (-1) \cdot a &= 1 \cdot a + (-1) \cdot a && \text{(definition of 1)} \\ &= (1 + -1) \cdot a && \text{(left distributivity)} \\ &= 0 \cdot a && \text{(definition of add inverse)} \\ &= 0 && \text{(From (a))} \end{aligned}$$

For (c), note that by right distributivity,

$$\begin{aligned} (-1) \cdot a + a &= (-1) \cdot a + 1 \cdot a && \text{(definition of 1)} \\ &= (-1 + 1) \cdot a && \text{(right distributivity)} \\ &= a \cdot 0 && \text{(definition of add inverse)} \\ &= 0 && \text{(From (a))} \end{aligned}$$

Therefore,

$$\begin{aligned} (-a)(-b) &= (-1 \cdot a)(-1 \cdot b) && \text{(from (b))} \\ &= -1 \cdot (a \cdot -1) \cdot b && \text{(associativity)} \\ &= -1 \cdot -a \cdot b && \text{(from (b))} \\ &= -1 \cdot -1 \cdot a \cdot b && \text{(from (b))} \\ &= (-1 \cdot -1) \cdot ab && \text{(associativity)} \\ &= 1ab && \text{(shown below)} \\ &= ab && \text{(definition of identity)} \end{aligned}$$

where $(-1)(-1) = 1$ since by (b), $(-1)(-1) = -(-1)$. We know that $-(-1)$ is an additive inverse for -1 and so is 1. Since the multiplicative identity is unique in a ring, $-(-1) = 1$. We show uniqueness for (d). Let us have $1 \neq 1'$. Then by definition of identity,

$$1 = 11' = 1'1 = 1' \tag{222}$$

which is a contradiction.

Exercise 5.26 (Shifrin 1.4.10)

1. Prove that the multiplicative inverse of a unit a in a ring R is unique. That is, if $ab = ba = 1$ and $ac = ca = 1$, then $b = c$. (You will need to use associativity of multiplication in R.)
2. Indeed, more is true. If $a \in R$ and there exist $b, c \in R$ so that $ab = 1$ and $ca = 1$, prove that $b = c$ and thus that a is a unit.

Solution. For (a), we see that

$$c = 1c = (ab)c = (ba)c = b(ac) = b(ca) = b1 = b \tag{223}$$

For (b), we have

$$b = 1b = (ca)b = c(ab) = c1 = c \tag{224}$$

Exercise 5.27 (Shifrin 1.4.13)

Let p be a prime number. Use the fact that $\mathbb{Z}_p$ is a field to prove that $(p-1)! \equiv -1 \pmod{p}$. (Hint: Pair elements of $\mathbb{Z}_p$ with their multiplicative inverses; cf. Exercise 1.3.12.).

Solution. For $p = 2$, the result is trivial. Now let $p > 2$ be a prime. Then since $\mathbb{F}$ is a field, every element $a \in \mathbb{F}$ contains a multiplicative inverse a^{-1}. We claim that the only values for which $a = a^{-1}$ is $1, p-1$. Assume that $a = a^{-1}$. Then

$$a^2 = 1 \implies p|(a^2 - 1) \implies p|(a+1)(a-1) \tag{225}$$

and since p is prime, it must be the case that $p|a+1 \iff a \equiv -1 \pmod{p}$ or $p|a-1 \iff a \equiv 1 \pmod{p}$. Therefore, we are left to consider the $(p-3)$ elements: $2, \dots, p-2$. Since inverses are unique and the inverses of inverses is the original element, we can partition these $p-2$ elements into $(p-3)/2$ pairs.[a] Let's call the set of pairs $K = \{(a, b)\}$ where $b = a^{-1}$. Therefore, by commutativity and associativity we have

$$(p-1)! \equiv (1)(p-1) \prod_{(a,b)\in K} ab \equiv -1 \cdot \prod_{(a,b)\in K} 1 \equiv -1 \pmod{p}. \tag{226}$$

[a] Since $p \neq 2$, p is odd and therefore $p-3$ is even.

Exercise 5.28 (Shifrin 2.3.2.a/b/c)

Recall that the conjugate of the complex number $z = a + bi$ is defined to be $\bar{z} = a - bi$. Prove the following properties of the conjugate:

1. $\overline{z + w} = \bar{z} + \bar{w}$
2. $\overline{zw} = \bar{z}\bar{w}$
3. $\bar{z} = z \iff z \in \mathbb{R}$ and $\bar{z} = -z \iff iz \in \mathbb{R}$
4. If $z = r(\cos\theta + i\sin\theta)$, then $\bar{z} = r(\cos\theta - i\sin\theta)$

Solution. Let $z = a + bi, w = c + di$. For (a),

$$\overline{z+w} = \overline{(a+c)+(b+d)i} = (a+c)-(b+d)i = a+c-bi-di = (a-bi)+(c-di) = \overline{z}+\overline{w} \tag{227}$$

For (b),

$$\overline{zw} = \overline{(ac-bd)+(ad+bc)i} = (ac-bd)-(ad+bc)i = ac-bd-adi-bci = (a-bi)(c-di) = \bar{z}\bar{w} \tag{228}$$

For (c), consider

$$\begin{aligned}
\overline{z} = z &\iff a + bi = a - bi && (229)\\
&\iff bi = -bi && (230)\\
&\iff 2bi = 0 && (231)\\
&\iff b = 0 && \text{(field has no 0 divisors)}
\end{aligned}$$

Therefore, $z = a \in \mathbb{R}$.

$$\begin{aligned}
\overline{z} = -z &\iff a - bi = -a - bi && (232)\\
&\iff a = -a && (233)\\
&\iff 2a = 0 && (234)\\
&\iff a = 0 && \text{(field has no 0 divisors.)}
\end{aligned}$$

Therefore, $z = bi \implies iz = -b \in \mathbb{R}$.

Exercise 5.29 (Shifrin 2.3.3.a/b/c)

Recall that the modulus of the complex number $z = a + bi$ is defined to be $|z| = \sqrt{a^2 + b^2}$. Prove the following properties of the modulus:

1. $|zw| = |z||w|$
2. $|\bar{z}| = |z|$
3. $|z|^2 = z\bar{z}$
4. $|z + w| \leq |z| + |w|$ (This is called the triangle inequality; why?)

Solution. Let $z = a + bi$ and $w = c + di$. For (a),

$$\begin{aligned}
|zw| &= |(ac - bd) + (ad + bc)i|\\
&= \sqrt{(ac - bd)^2 + (ad + bc)^2}\\
&= \sqrt{a^2c^2 - 2abcd + b^2d^2 + a^2d^2 + 2abcd + b^2c^2}\\
&= \sqrt{(a^2 + b^2)(c^2 + d^2)}\\
&= \sqrt{a^2 + b^2}\sqrt{c^2 + d^2}\\
&= |z||w|
\end{aligned}$$

For (b), if $z = a + bi$, then $\bar{z} = a - bi$, so:

$$|\bar{z}| = \sqrt{a^2 + (-b)^2} = \sqrt{a^2 + b^2} = |z| \tag{235}$$

For (c),

$$\begin{aligned}
z\bar{z} &= (a + bi)(a - bi)\\
&= a^2 + b^2\\
&= |z|^2
\end{aligned}$$

Exercise 5.30 (Shifrin 3.1.2.c/d)

Find the greatest common divisors $d(x)$ of the following polynomials $f(x), g(x) \in F[x]$, and express $d(x)$ as $s(x)f(x) + t(x)g(x)$ for appropriate $s(x), t(x) \in F[x]$:

1. $f(x) = x^3 - 1$, $g(x) = x^4 + x^3 - x^2 - 2x - 2$, $F = \mathbb{Q}$
2. $f(x) = x^2 + (1 - \sqrt{2})x - \sqrt{2}$, $g(x) = x^2 - 2$, $F = \mathbb{R}$
3. $f(x) = x^2 + 1$, $g(x) = x^2 - i + 2$, $F = \mathbb{C}$
4. $f(x) = x^2 + 2x + 2$, $g(x) = x^2 + 1$, $F = \mathbb{Q}$
5. $f(x) = x^2 + 2x + 2$, $g(x) = x^2 + 1$, $F = \mathbb{C}$

Solution. For (c), the gcd is 1, with

$$-\frac{1}{1-i}(x^2+1) + \frac{1}{1-i}(x^2 - i + 2) = \frac{1}{1-i}(x^2 - i + 2 - x^2 - 1) = \frac{1}{1-i}(1-i) = 1 \tag{236}$$

where $1/(1-i) = (1+i)/2$. For (d), the gcd is 1, with

$$\frac{1}{5}(2x+3)(x^2+1) + \frac{1}{5}(1-2x)(x^2+2x+2) \tag{237}$$

$$= \frac{1}{5}(2x^3 + 3x^2 + 2x + 3) + \frac{1}{5}(-2x^3 - 3x^2 - 2x + 2) = 1 \tag{238}$$

Exercise 5.31 (Shifrin 3.1.6)

Prove that if F is a field, $f(x) \in F[x]$, and $\deg(f(x)) = n$, then $f(x)$ has at most n roots in F.

Solution. We start when $n = 1$. Then $f(x) = mx + b$ and we claim that the only root is $x = -b/m$ since we can solve for $0 = mx + b$ with the field operations, which leads to a unique solution. This implies by corr 1.5 that $(x + b/m)$ is the only factor of f. Now suppose this holds true for some degree $n - 1$ and let us have a degree n polynomial f. Assume that some c is a root of f (if there exists no c, then we are trivially done), which means $(x - c)$ is a factor of f, and we can write

$$f(x) = (x - c)\, g(x) \tag{239}$$

for some polynomial $g(x)$ of degree $n - 1$. By our inductive hypothesis, $g(x)$ must have at most $n - 1$ roots, and so f has at most n roots.

Exercise 5.32 (Shifrin 3.1.8)

Let F be a field. Prove that if $f(x) \in F[x]$ is a polynomial of degree 2 or 3, then $f(x)$ is irreducible in $F[x]$ if and only if $f(x)$ has no root in F.

Solution. We prove bidirectionally.

1. ($\rightarrow$). Let f be irreducible. Then it cannot be factored into polynomials $p(x)q(x)$ where $\deg(p) + \deg(q) = n$. Note that two positive integers adding up to 2 or 3 means that at least one of the integers must be 1, by the pigeonhole principle. This means that f irreducible is equivalent to saying that f does not have linear factors of form $(x - c)$, which by corollary 1.5 implies that there exists no root c for $f(x)$.
2. ($\leftarrow$). Let f have no root in F. Then by corollary 1.5 there exists no linear factors $(x - c)$. By the same pigeonhole principle argument, we know that having a linear factor for degree 2 or 3 polynomials is equivalent to having (general) factors, and so f has no factors. Therefore f is irreducible.

Exercise 5.33 (Shifrin 3.1.13)

List all the irreducible polynomials in $\mathbb{Z}_2[x]$ of degree ≤ 4. Factor $f(x) = x^7 + 1$ as a product of irreducible polynomials in $\mathbb{Z}_2[x]$.

Solution. Listed by degree.
1. 1: $x, x+1$.
2. 2: $x^2 + x + 1$.
3. 3: $x^3 + x^2 + 1, x^3 + x + 1$.
4. 4: $x^4 + x + 1, x^4 + x^3 + 1, x^4 + x^3 + x^2 + x + 1$.

We have

$$\begin{aligned} x^7 + 1 &= (x+1)(x^6 + x^5 + x^4 + x^3 + x^2 + x + 1) && (240)\\ &= (x+1)(x^3 + x + 1)(x^3 + x^2 + 1) && (241)\end{aligned}$$

Exercise 5.34 (Shifrin 3.2.2.b/c)

Prove that
1. $\mathbb{Q}[\sqrt{2}, i] = \mathbb{Q}[\sqrt{2} + i]$, but $\mathbb{Q}[\sqrt{2}i] \subsetneq \mathbb{Q}[\sqrt{2}, i]$
2. $\mathbb{Q}[\sqrt{2}, \sqrt{3}] = \mathbb{Q}[\sqrt{2} + \sqrt{3}]$, but $\mathbb{Q}[\sqrt{6}] \subsetneq \mathbb{Q}[\sqrt{2}, \sqrt{3}]$
3. $\mathbb{Q}[\sqrt[3]{2} + i] = \mathbb{Q}[\sqrt[3]{2}, i]$; what about $\mathbb{Q}[\sqrt[3]{2}i] \subset \mathbb{Q}[\sqrt[3]{2}, i]$?

Solution. From Shifrin, I use the fact that $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$, and the same proof immediately shows that $\mathbb{Q}[\sqrt{3}] = \{a + b\sqrt{3} \mid a, b \in \mathbb{Q}\}$ along with that for $\mathbb{Q}[\sqrt{6}]$. As for $\mathbb{Q}[\sqrt{2}, \sqrt{3}]$, I also follow the same logic to show

$$\begin{aligned} \mathbb{Q}[\sqrt{2}, \sqrt{3}] &= \mathbb{Q}[\sqrt{2}][\sqrt{3}] && (242)\\ &= \{\alpha + \beta\sqrt{3} \mid a, b \in \mathbb{Q}[\sqrt{2}]\} && (243)\\ &= \{(a + b\sqrt{2}) + (c + d\sqrt{2})\sqrt{3} \mid a, b, c, d \in \mathbb{Q}\} && (244)\\ &= \{a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6} \mid a, b, c, d \in \mathbb{Q}\} && (245)\end{aligned}$$

Where $\sqrt{2} \times \sqrt{3} = \sqrt{2 \times 3} = \sqrt{6}$ follows from the definition of nth roots plus associativity on the reals. For (b), we prove bidirectionally.
1. $\mathbb{Q}[\sqrt{2} + \sqrt{3}] \subset \mathbb{Q}[\sqrt{2}, \sqrt{3}]$. Consider $y \in \mathbb{Q}[\sqrt{2} + \sqrt{3}]$. Then there exists $p \in \mathbb{Q}[x]$ s.t.
$$y = p(\sqrt{2} + \sqrt{3}) = a_n(\sqrt{2} + \sqrt{3})^n + \ldots + a_1(\sqrt{2} + \sqrt{3}) + a_0 \quad (246)$$
where the terms can be expanded an rearranged to the form $a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6} \in \mathbb{Q}[\sqrt{2}, \sqrt{3}]$.
2. $\mathbb{Q}[\sqrt{2}, \sqrt{3}] \subset \mathbb{Q}[\sqrt{2} + \sqrt{3}]$. Consider $\sqrt{2} + \sqrt{3} \in \mathbb{Q}[\sqrt{2} + \sqrt{3}]$. Since it is a field and $\sqrt{2} + \sqrt{3}$ is a unit, by rationalizing the denominator, we can get
$$(\sqrt{2} + \sqrt{3})^{-1} = \frac{\sqrt{2} - \sqrt{3}}{2 - 3} = \sqrt{3} - \sqrt{2} \in \mathbb{Q}[\sqrt{2} + \sqrt{3}] \quad (247)$$
Therefore by adding and subtracting the two elements, we have $\sqrt{2}, \sqrt{3} \in \mathbb{Q}[\sqrt{2} + \sqrt{3}] \implies \sqrt{6} \in \mathbb{Q}[\sqrt{2} + \sqrt{3}]$. Since $\mathbb{Q} \subset \mathbb{Q}[\sqrt{2} + \sqrt{3}]$, from the ring properties all elements of the form $a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6} \in \mathbb{Q}[\sqrt{2} + \sqrt{3}]$.

For the second part, I claim that $\sqrt{2} \notin \mathbb{Q}[\sqrt{6}]$. Assuming it is, we have $\sqrt{2} = a + b\sqrt{6} \implies 2 = a^2 + 6b^2 + 2ab\sqrt{6}$. So $a = 0$ or $b = 0$. If $a = 0$, then $b^2 = 1/3 \implies b = 1/\sqrt{3}$ which contradicts that b is rational. If $b = 0$, then $a^2 = 2 \implies a = \sqrt{2}$ which contradicts that a is rational. Note that

$\mathbb{Q}[\sqrt[3]{2}] = \{a + b\sqrt[3]{2} + c\sqrt[3]{4}\}$, and so

$$\begin{aligned}\mathbb{Q}[\sqrt[3]{2}, i] &= \mathbb{Q}[\sqrt[3]{2}][i] && (248)\\ &= \{\alpha + \beta i \mid \alpha, \beta \in \mathbb{Q}[\sqrt[3]{2}]\} && (249)\\ &= \{(a + b\sqrt[3]{2} + c\sqrt[3]{4}) + (d + e\sqrt[3]{2} + f\sqrt[3]{4})i \mid a, b, c, d, e, f \in \mathbb{Q}\} && (250)\\ &= \{a + b\sqrt[3]{2} + c\sqrt[3]{4} + di + e\sqrt[3]{2}i + f\sqrt[3]{4}i \mid a, b, c, d, e, f \in \mathbb{Q}\} && (251)\end{aligned}$$

We prove bidirectionally.

1. $\mathbb{Q}[\sqrt[3]{2} + i] \subset \mathbb{Q}[\sqrt[3]{2}, i]$. Consider $y \in \mathbb{Q}[\sqrt[3]{2} + i]$. Then there exists a $p \in \mathbb{Q}[x]$ s.t.

$$y = p(\sqrt[3]{2} + i) = a_n(\sqrt[3]{2} + i)^n + \ldots + a_1(\sqrt[3]{2} + i) + a_0 \tag{252}$$

   Then we can expand and rearrange the terms to be of the form

$$a + b\sqrt[3]{2} + c\sqrt[3]{4} + di + ei\sqrt[3]{2} + fi\sqrt[3]{4} \in \mathbb{Q}[\sqrt[3]{2}, i] \tag{253}$$

2. $\mathbb{Q}[\sqrt[3]{2}, i] \subset \mathbb{Q}[\sqrt[3]{2} + i]$. Consider $\alpha = \sqrt[3]{2} + i \in \mathbb{Q}[\sqrt[3]{2} + i]$. Then $(\alpha - i)^3 = 2$. Therefore

$$\begin{aligned}\alpha^3 - 3\alpha^2 i - 3\alpha + i = 2 &\implies i(1 - 3\alpha^2) = 2 + 3\alpha - \alpha^3 && (254)\\ &\implies i = \frac{2 + 3\alpha - \alpha^3}{1 - 3\alpha^2} \in \mathbb{Q}[\sqrt[3]{2} + i] && (255)\end{aligned}$$

   Therefore $\sqrt[3]{2} = \alpha - i \in \mathbb{Q}[\sqrt[3]{2} + i]$, which allows us add all combinations $\{1, \sqrt[3]{2}, \sqrt[3]{4}, i, \sqrt[3]{2}i, \sqrt[3]{4}i\}$ into our basis.

Exercise 5.35 (Shifrin 3.2.6.b/c/d/g)

Suppose $\alpha \in \mathbb{C}$ is a root of the given irreducible polynomial $f(x) \in \mathbb{Q}[x]$. Find the multiplicative inverse of $\beta \in \mathbb{Q}[\alpha]$.

1. $f(x) = x^2 + 3x - 3$, $\beta = \alpha - 1$
2. $f(x) = x^3 + x^2 - 2x - 1$, $\beta = \alpha + 1$
3. $f(x) = x^3 + x^2 + 2x + 1$, $\beta = \alpha^2 + 1$
4. $f(x) = x^3 - 2$, $\beta = \alpha + 1$
5. $f(x) = x^3 + x^2 - x + 1$, $\beta = \alpha + 2$
6. $f(x) = x^3 - 2$, $\beta = r + s\alpha + t\alpha^2$
7. $f(x) = x^4 + x^2 - 1$, $\beta = \alpha^3 + \alpha - 1$

Solution. For (b), using the Euclidean algorithm gives

$$(1)(x^3 + x^2 - 2x - 1) + (-x^2 + 2)(x + 1) = 1 \tag{256}$$

and substituting the root α gives $(-\alpha^2 + 2)(\alpha + 1) = 1$. So we have $\beta^{-1} = -\alpha^2 + 2$. For (c), doing the same thing gives

$$(-x)(x^3 + x^2 + 2x + 1) + (x^2 + x + 1)(x^2 + 1) = 1 \tag{257}$$

and substituting α gives $(\alpha^2 + \alpha + 1)(\alpha^2 + 1) = 1$, so $\beta^{-1} = \alpha^2 + \alpha + 1$. For (d), we have

$$(-\frac{1}{3})(x^3 - 2) + (\frac{1}{3}x^2 - \frac{1}{3}x + \frac{1}{3})(x + 1) = 1 \tag{258}$$

and so substituting α gives $(\frac{1}{3}\alpha^2 - \frac{1}{3}\alpha + \frac{1}{3})(\alpha + 1) = 1$, so $\beta^{-1} = \frac{1}{3}\alpha^2 - \frac{1}{3}\alpha + \frac{1}{3}$. For (g), we have

$$(-x^2 - x - 2)(x^4 + x^2 - 1) + (x^3 + x^2 + 2x + 1)(x^3 + x - 1) = 1 \tag{259}$$

and so substituting α gives $(\alpha^3 + \alpha^2 + 2\alpha + 1)(\alpha^3 + \alpha - 1) = 1$, and so $\beta^{-1} = \alpha^3 + \alpha^2 + 2\alpha + 1$.

Exercise 5.36 (Shifrin 3.2.7)

Let $f(x) \in \mathbb{R}[x]$.

1. Prove that the complex roots of $f(x)$ come in "conjugate pairs"; i.e., $\alpha \in \mathbb{C}$ is a root of $f(x)$ if and only if $\overline{\alpha}$ is also a root.
2. Prove that the only irreducible polynomials in $\mathbb{R}[x]$ are linear polynomials and quadratic polynomials $ax^2 + bx + c$ with $b^2 - 4ac < 0$.

Solution. Listed.

1. If $\alpha \in \mathbb{C}$ is a root of f, then

$$0 = f(\alpha) = a_n\alpha^n + \ldots + a_1\alpha + a_0 \tag{260}$$

for $a_i \in \mathbb{R}$. Since

$$\begin{aligned} 0 = \overline{0} &= \overline{f(\alpha)} && (261)\\ &= \overline{a_n\alpha^n + \ldots + a_1\alpha + a_0} && (262)\\ &= \overline{a_n}\overline{\alpha^n} + \ldots + \overline{a_1\alpha} + \overline{a_0} && (263)\\ &= a_n\overline{\alpha}^n + \ldots + a_1\overline{\alpha} + a_0 && (264)\\ &= p(\overline{\alpha}) && (265) \end{aligned}$$

we can see that $\overline{\alpha} \in \mathbb{C}$ is immediately a root as well. Since $\overline{\overline{\alpha}} = \alpha$, the converse is immediately proven.

2. Linear polynomials in $F[x]$ for a given field are trivially irreducible (since multiplying polynomials increases the degree of the product as there are no zero divisors in a field). Perhaps without Theorem 4.1, we can assume that a real quadratic polynomial $p(x) = ax^2 + bx + c$ is reducible, which is equivalent to

$$p(x) = (dx + e)(fx + g) = dfx^2 + (dg + ef)x + eg \tag{266}$$

For $d, e, f, g \in \mathbb{R}$, and evaluating $b^2 - 4ac = (dg + ef)^2 - 4dfeg = (dg - ef)^2 \geq 0$ since this is a squared term of a real number. So we have proved that if it is quadratic and reducible, then the discriminant ≥ 0. To prove the other way, we assume that it is not reducible, i.e. there exists some complex root α from the fundamental theorem of algebra. Then from (1), we know that $\overline{\alpha}$ must also be a complex conjugate. Then this is reducible in $\mathbb{C}$ as

$$p(x) = a(x - \alpha)(x - \overline{\alpha}) \tag{267}$$

for some constant factor a. Letting $\alpha = d + ei$ for $d, e \in \mathbb{R}$, expanding it gives us

$$\begin{aligned} p(x) &= a\big(x^2 - (\alpha + \overline{\alpha})x + \alpha\overline{\alpha}\big) && (268)\\ &= ax^2 + -2adx + a(d^2 + e^2) && (269) \end{aligned}$$

and evaluating the discriminant gives

$$4a^2d^2 - 4a^2(d^2 + e^2) = -4a^2e^2 < 0 \tag{270}$$

and we are done. For higher degree polynomials, we can proceed by taking a complex root (which is guaranteed to exist by fundamental theorem of algebra). If it contains an imaginary term, then its conjugate is also a root, and we factor out the quadratic. If it is real, then we can factor out the linear term. We can keep going this until we hit our base cases of a quadratic or linear term.

Exercise 5.37 (Shifrin 3.2.13)

Let K be a field extension of F, and suppose $\alpha, \beta \in K$. Show that $(F[\alpha])[\beta] = (F[\beta])[\alpha]$, so that $F[\alpha, \beta]$ makes good sense.
(Remark: One way to do this is to think about the ring of polynomials in two variables. The other way is just to show directly that every element of one ring belongs to the other.)

Solution. Let $y \in (F[\alpha])[\beta]$. Then there exists a polynomial $p \in (F[\alpha])[x]$ s.t.

$$y = p(\beta) = b_n\beta^n + \ldots + b_1\beta + b_0 = \sum_{i=0}^{n} b_i\beta^i \tag{271}$$

for $b_i \in F[\alpha]$. But since $b_i \in F[\alpha]$, there exists a polynomial $q_i \in F[x]$ s.t. (omitting the subscript i for clarity)

$$b_i = q_i(\alpha) = a_{n_i}\alpha^n + \ldots + a_1\alpha + a_0 = \sum_{j=0}^{n_i} a_j\alpha^j \tag{272}$$

for $a_j \in F$. Substituting each b_i in gives

$$y = \sum_{i=0}^{n}\left(\sum_{j=0}^{n_i} a_j\alpha^j\right)\beta^i = \sum_{i=0}^{n}\sum_{j=0}^{n_i} a_j\alpha^j\beta^i \tag{273}$$

With the same logic, every element of $(F[\beta])[\alpha]$ can be written as

$$y = \sum_{i=0}^{n}\left(\sum_{j=0}^{n_i} a_j\beta^j\right)\alpha^i = \sum_{i=0}^{n}\sum_{j=0}^{n_i} a_j\alpha^i\beta^j \tag{274}$$

Note that since $F[\alpha]$ is a vector space spanned by $\{1, \ldots, \alpha^{n-1}\}$, and $F[\beta]$ is a also a vector space spanned by $\{1, \ldots, \beta^{m-1}\}$ for some m, the two spaces above are spanned by all products $\{\alpha^i\beta^j\}_{i<n, j<m}$, and they are the same set.

Exercise 5.38 (Shifrin 3.3.2.a/d/e/g)

Decide which of the following polynomials are irreducible in $\mathbb{Q}[x]$.

a $f(x) = x^3 + 4x^2 - 3x + 5$
1. $f(x) = 4x^4 - 6x^2 + 6x - 12$
2. $f(x) = x^3 + x^2 + x + 1$
d $f(x) = x^4 - 180$
e $f(x) = x^4 + x^2 - 6$
3. $f(x) = x^4 - 2x^3 + x^2 + 1$
g $f(x) = x^3 + 17x + 36$
4. $f(x) = x^4 + x + 1$
5. $f(x) = x^5 + x^3 + x^2 + 1$
6. $f(x) = x^5 + x^3 + x + 1$

Solution. For (a), by the rational root theorem the rational roots, if any, must be in the set $\{\pm 1, \pm 5\}$. Calculating them gives $f(x) = 7, 11, 215, -5$. Since this is third degree, no linear factors means that it is irreducible, so f is irreducible.
For (d), by the Eisenstein's criterion with $p = 5$ this polynomial is irreducible.
For (e), the rational root theorem states that the rational roots must be in $\{\pm 1, \pm 2, \pm 3, \pm 6\}$. This polynomial is clearly even, so it suffices to check the positive candidates. This gives $-4, 14, 84, 1326$.

Therefore if it is reducible, by Gauss's lemma it must be of the form

$$(ax^2 + bx + c)(dx^2 + ex + f) \tag{275}$$

for integer coefficients. $a = d = 1$ is trivial ($-1, -1$ is also possible but constant factors don't matter). Expanding this gives

$$x^4 + (b + e)x^3 + (c + f + be)x^2 + (bf + ce)x + cf = x^4 + x^2 - 6 \tag{276}$$

The coefficients of x^3 tell us that $e = -b$, which means that for the coefficents of x, $bf + ce = bf - bc = 0 \implies f = c$. So $c^2 = -6$, which has no solution. Therefore f is irreducible.
For (g), we must check rational roots of $\{\pm 1, \pm 2, \pm 3, \pm 4, \pm 6, \pm 9, \pm 12, \pm 18, \pm 36\}$. Since this polynomial is monotonically increasing, with $f(-2) = -6$ and $f(0) = 36$. It only suffices to check $x = -1$, which gives $f(-1) = 18$. Therefore there are no linear factors. Since this is third degree, no linear factors means that it is irreducible, so f is irreducible.

Exercise 5.39 (Shifrin 3.3.4)

Show that each of the following polynomials has no rational root:
1. $x^{200} - x^{41} + 4x + 1$
2. $x^8 - 54$
3. $x^{2k} + 3x^{k+1} - 12$, $k \geq 1$

Solution. Listed.
1. By the rational root theorem, the only possible rational roots are ± 1. Solving for both of these values gives

$$f(1) = 1 - 1 + 4 + 1 = 5 \tag{277}$$
$$f(-1) = 1 + 1 - 4 + 1 = -1 \tag{278}$$

   Therefore there are no rational roots.
2. The only possible rational roots are $\pm 1, \pm 2, \pm 3, \pm 6, \pm 9, \pm 18, \pm 27, \pm 54$. But this polynomial is even, so it suffices to check the positive roots. $f(1) = -53$, $f(2) = 256 - 54 = 202$, and any greater inputs will increase the output since f is monotonic in $\mathbb{Z}^+$. Therefore f has no rational roots.
3. By Eisenstein's criterion with $p = 3$, this polynomial is irreducible and therefore has no rational roots.

Exercise 5.40 (Shifrin 3.3.6)

Listed.
1. Prove that $f(x) \in \mathbb{Z}_2[x]$ has $x + 1$ as a factor if and only if it has an even number of nonzero coefficients.
2. List the irreducible polynomials in $\mathbb{Z}_2[x]$ of degrees $2, 3, 4$, and 5.

Solution. Listed. Since $f(x)$ has $x + 1$ as a factor iff

$$f(1) = a_n 1^n + \ldots + a_1 1^1 + a_0 = a_n + \ldots + a_1 + a_0 = 0 \tag{279}$$

where each $a_i \in \{0, 1\}$. Therefore, this is equivalent to saying that there are an even number of 1's (nonzero coefficients), which sum to 0 mod 2. Therefore, the irreducible polynomials should at least have a constant coefficient of 1 (so we can't factor x) and should have odd number of terms (so that

we can't factor $x+1$). This will guarantee that $f(0) = f(1) = 1$.

1. Degree 2: x^2+x+1 is the only candidate and indeed is an irreducible polynomial.
2. Degree 3: x^3+x^2+1, x^3+x+1 and indeed $f(0) = f(1) = 1$. Since it's only degree 3 we don't need to check irreducibility into 2 terms of both degree at least 2.
3. Degree 4: $x^4+x^3+x^2+x+1$, x^4+x^3+1, x^4+x^2+1, x^4+x+1 are candidates. However we need to check that they cannot be factored into two irreducible quadratic polynomials. The only possible such factorization is

$$(x^2+x+1)(x^2+x+1) = x^4+x^2+1 \tag{280}$$

and so the irreducible polynomials are $x^4+x^3+x^2+x+1$, x^4+x^3+1, x^4+x+1.

4. Degree 5: x^5+x^4+1, x^5+x^3+1, x^5+x^2+1, x^5+x+1, $x^5+x^4+x^3+x^2+1$, $x^5+x^4+x^3+x+1$, $x^5+x^4+x^2+x+1$, $x^5+x^3+x^2+x+1$ are the possible candidates. But we need to check that it is not factorable into an irreducible quadratic and cubic. The three candidates are

$$(x^2+x+1)(x^3+x^2+1) = x^5+x+1 \tag{281}$$

$$(x^2+x+1)(x^3+x+1) = x^5+x^4+1 \tag{282}$$

and so the irreducible polynomials are x^5+x^3+1, x^5+x^2+1, $x^5+x^4+x^3+x^2+1$, $x^5+x^4+x^3+x+1$, $x^5+x^4+x^2+x+1$, $x^5+x^3+x^2+x+1$.

Exercise 5.41 (Shifrin 3.3.7)

Prove that for any prime number p, $f(x) = x^{p-1}+x^{p-2}+\cdots+x+1$ is irreducible in $\mathbb{Q}[x]$.

Solution. We can use the identity

$$f(x) = x^{p-1}+x^{p-2}+\cdots+x+1 = \frac{x^p-1}{x-1} \tag{283}$$

Therefore,

$$f(x+1) = \frac{(x+1)^p-1}{(x+1)-1} = \frac{1}{x}\left\{\left(\sum_{k=0}^{p}\binom{p}{k}x^k\right)-1\right\} \tag{284}$$

$$= \frac{1}{x}\sum_{k=1}^{p}\binom{p}{k}x^k = \sum_{k=1}^{p}\binom{p}{k}x^{k-1} \tag{285}$$

Focusing on the coefficients, the leading coefficient is $\binom{p}{p} = 1$, and the rest of the coefficients are divisible by p. The constant coefficient is $\binom{p}{1} = p$, which is not divisible by p^2. By Eisenstein's criterion, $f(x+1)$ is irreducible $\implies$ $f(x)$ is irreducible. To justify the final step, assume that $f(x)$ is reducible. Then $f(x) = g(x)h(x)$ for positive degree polynomials g, h. Then by substituting $x+1$, we have that $f(x+1) = g(x+1)h(x+1)$, which means that $f(x+1)$ is irreducible.

Exercise 5.42 (Shifrin 4.1.3)

(a) Prove that if $I \subset R$ is an ideal and $1 \in I$, then $I = R$.
(b) Prove that $a \in R$ is a unit if and only if $\langle a \rangle = R$.
(c) Prove that the only ideals in a (commutative) ring R are $\langle 0 \rangle$ and R if and only if R is a field.

Solution. Listed.
(a) If $1 \in I$, then for every $r \in R$, we must have $r1 = r \in I$. Therefore $I = R$.
(b) If $a \in R$ is a unit, then $a^{-1} \in R$, and so for every $r \in R$, $ra^{-1} \in R$. Therefore, $\langle a \rangle$ must contain all elements of form $ra^{-1}a = r$, which is precisely R. Now assume that a is not a unit, and so there exists no $a^{-1} \in R$. Therefore, $\langle a \rangle$, which consists of all ra for $r \in R$, cannot contain 1 since $r \neq a^{-1}$, and so $\langle a \rangle \neq R$.
(c) For the forwards implication, assume that R is not a field. Then there exists some $a \neq 0$ that is not a unit, and taking $\langle a \rangle$ gives us an ideal that—from (b)—is not R. For the backward implication we know that $\langle 0 \rangle$ is an ideal. Now assume that there exists another ideal I containing $a \neq 0$. Since R is a field, a is a unit, and so by (b) $R = \langle a \rangle \subset I \subset R \implies I = R$.

Exercise 5.43 (Shifrin 4.1.4.a/b/c)

Find all the ideals in the following rings:
(a) $\mathbb{Z}$
(b) $\mathbb{Z}_7$
(c) $\mathbb{Z}_6$
(d) $\mathbb{Z}_{12}$
(e) $\mathbb{Z}_{36}$
(f) $\mathbb{Q}$
(g) $\mathbb{Z}[i]$ (see Exercise 2.3.18)

Solution. Listed.
(a) All sets of form $\{kz \in \mathbb{Z} \mid z \in \mathbb{Z}\}$ for all $k \in \mathbb{Z}$.
(b) Only $\{0\}$ and $\mathbb{Z}_7$ is an ideal.
(c) We have $\{0\}, \{0, 2, 4\}, \{0, 3\}, \mathbb{Z}_6$.

Exercise 5.44 (Shifrin 4.1.5)

(a) Let $I = \langle f(x) \rangle$, $J = \langle g(x) \rangle$ be ideals in $F[x]$. Prove that $I \subset J \Leftrightarrow g(x) | f(x)$.
(b) List all the ideals of $\mathbb{Q}[x]$ containing the element $f(x) = (x^2 + x - 1)^3(x - 3)^2$.

Solution. For (a), we prove bidirectionally.
1. ($\rightarrow$). Since $f(x) \in \langle f(x) \rangle \implies f(x) \in \langle g(x) \rangle$, this means that $f(x) = r(x)g(x)$ for some $r(x) \in F[x$. Therefore $g(x) \mid f(x)$.
2. ($\leftarrow$). Given that $g(x) \mid f(x)$, let us take some $f_1(x) \in I$. Then it is of the form $f_1(x) = r(x)f(x)$ for some $r(x) \in F[x]$. But since $g(x) \mid f(x)$, $f(x) = h(x)g(x)$ for some $h(x) \in F[x]$. Therefore $f_1(x) = r(x)h(x)g(x) = (rh)(x)g(x)$, where $(rh)(x) \in F[x]$, and so $f_1(x) \in J$.

For (b), we can use the logic from (a) to find all the factors of $f(x)$, which generate all sup-ideals of $\langle f(x) \rangle$, which is the minimal ideal containing $f(x)$.
1. $g(x) = 1 \implies \langle 1 \rangle = F[x]$
2. $g(x) = x^2 + x - 1 \implies \langle x^2 + x - 1 \rangle$
3. $g(x) = (x^2 + x - 1)^2 \implies \langle (x^2 + x - 1)^2 \rangle$
4. $g(x) = (x^2 + x - 1)^3 \implies \langle (x^2 + x - 1)^3 \rangle$
5. $g(x) = x - 3 \implies \langle x - 3 \rangle$
6. $g(x) = (x^2 + x - 1)(x - 3) \implies \langle (x^2 + x - 1)(x - 3) \rangle$
7. $g(x) = (x^2 + x - 1)^2(x - 3) \implies \langle (x^2 + x - 1)^2(x - 3) \rangle$
8. $g(x) = (x^2 + x - 1)^3(x - 3) \implies \langle (x^2 + x - 1)^3(x - 3) \rangle$
9. $g(x) = (x - 3)^2 \implies \langle (x - 3)^2 \rangle$
10. $g(x) = (x^2 + x - 1)(x - 3)^2 \implies \langle (x^2 + x - 1)(x - 3)^2 \rangle$

11. $g(x) = (x^2+x-1)^2(x-3)^2 \implies \langle (x^2+x-1)^2(x-3)^2 \rangle$
12. $g(x) = (x^2+x-1)^3(x-3)^2 \implies \langle (x^2+x-1)^3(x-3)^2 \rangle$

Exercise 5.45 (Shifrin 4.1.14.a/b)

Mimicking Example 5(c), give the addition and multiplication tables of
(a) $\mathbb{Z}_2[x]/\langle x^2+x \rangle$
(b) $\mathbb{Z}_3[x]/\langle x^2+x-1 \rangle$
(c) $\mathbb{Z}_2[x]/\langle x^3+x+1 \rangle$
In each case, is the quotient ring an integral domain? a field?

Solution. For (a), note that the quotient allows us to state that $x^2 \equiv x \pmod I$, and therefore every polynomial in $\mathbb{Z}_2[x]/\langle x^2+x \rangle$ is equivalent to a linear polynomial. Therefore, the elements in this quotient are $0, 1, x, x+1$. As you can see, this is not an integral domain (and hence not a field) since $x, x+1$ are zero divisors.

$+$	0	1	x	$x+1$
0	0	1	x	$x+1$
1	1	0	$x+1$	x
x	x	$x+1$	0	1
$x+1$	$x+1$	x	1	0

$\times$	0	1	x	$x+1$
0	0	0	0	0
1	0	1	x	$x+1$
x	0	x	x	0
$x+1$	0	$x+1$	0	$x+1$

Figure 12: Addition and multiplication tables for $\mathbb{Z}_2[x]/\langle x^2+x \rangle$.

For (b), note that the quotient allows us to state that $x^2 \equiv 2x+1 \pmod I$, and therefore every polynomial in $\mathbb{Z}_3[x]/\langle x^2+x-1 \rangle$ is equivalent to a linear polynomial. Therefore, the elements in this quotient are $0, 1, 2, x, x+1, x+2, 2x, 2x+1, 2x+2$. This is indeed an integral domain since there are no zero divisors, and it is a field since every nonzero element is a unit (all rows/columns are filled with all elements of the set).

$+$	0	1	2	x	$x+1$	$x+2$	$2x$	$2x+1$	$2x+2$
0	0	1	2	x	$x+1$	$x+2$	$2x$	$2x+1$	$2x+2$
1	1	2	0	$x+1$	$x+2$	x	$2x+1$	$2x+2$	$2x$
2	2	0	1	$x+2$	x	$x+1$	$2x+2$	$2x$	$2x+1$
x	x	$x+1$	$x+2$	$2x$	$2x+1$	$2x+2$	0	1	2
$x+1$	$x+1$	$x+2$	x	$2x+1$	$2x+2$	$2x$	1	2	0
$x+2$	$x+2$	x	$x+1$	$2x+2$	$2x$	$2x+1$	2	0	1
$2x$	$2x$	$2x+1$	$2x+2$	0	1	2	x	$x+1$	$x+2$
$2x+1$	$2x+1$	$2x+2$	$2x$	1	2	0	$x+1$	$x+2$	x
$2x+2$	$2x+2$	$2x$	$2x+1$	2	0	1	$x+2$	x	$x+1$

Figure 13: Addition table for $\mathbb{Z}_3[x]/\langle x^2+x-1 \rangle$.

$\times$	0	1	2	x	$x+1$	$x+2$	$2x$	$2x+1$	$2x+2$
0	0	0	0	0	0	0	0	0	0
1	0	1	2	x	$x+1$	$x+2$	$2x$	$2x+1$	$2x+2$
2	0	2	1	$2x$	$2x+2$	$2x+1$	x	$x+2$	$x+1$
x	0	x	$2x$	$2x+1$	1	$x+1$	$x+2$	$2x+2$	2
$x+1$	0	$x+1$	$2x+2$	1	$x+2$	$2x$	2	x	$2x+1$
$x+2$	0	$x+2$	$2x+1$	$x+1$	$2x$	2	$2x+2$	1	x
$2x$	0	$2x$	x	$x+2$	2	$2x+2$	$2x+1$	$x+1$	1
$2x+1$	0	$2x+1$	$x+2$	$2x+2$	x	1	$x+1$	2	$2x$
$2x+2$	0	$2x+2$	$x+1$	2	$2x+1$	x	1	$2x$	$x+2$

Figure 14: Multiplication table for $\mathbb{Z}_3[x]/\langle x^2+x-1\rangle$.

Exercise 5.46 (Shifrin 4.1.17)

Let R be a commutative ring and let $I, J \subset R$ be ideals. Define

$$I \cap J = \{a \in R : a \in I \text{ and } a \in J\}$$
$$I + J = \{a + b \in R : a \in I, b \in J\}.$$

(a) Prove that $I \cap J$ and $I + J$ are ideals.
(b) Suppose $R = \mathbb{Z}$ or $F[x]$, $I = \langle a\rangle$, and $J = \langle b\rangle$. Identify $I \cap J$ and $I + J$.
(c) Let $a_1, \dots, a_n \in R$. Prove that $\langle a_1, \dots, a_n\rangle = \langle a_1\rangle + \cdots + \langle a_n\rangle$.

Solution. For (a), we prove it in 5.11. For (b), the argument is equivalent for $\mathbb{Z}$ and $F[x]$. $I \cap J$ consists of all elements that are divisible by both a and b, so $I \cap J = \langle \text{lcm}(a, b)\rangle$. $I + J$ consists of all elements that are of form $ra + sb$, but this are all multiples of $\gcd(a, b)$ and so $I + J = \langle \gcd(a, b)\rangle$.
For (c), it suffices to prove $\langle a, b\rangle = \langle a\rangle + \langle b\rangle$.

1. $\langle a, b\rangle \subset \langle a\rangle + \langle b\rangle$. $x \in \langle a, b\rangle \implies x = r_a a + r_b b$ for $r_a, r_b \in R$. But $a \in \langle a\rangle, b \in \langle b\rangle \implies r_a a \in \langle a\rangle, r_b b \in \langle b\rangle$, and so $x \in \langle a\rangle + \langle b\rangle$.
2. $\langle a, b\rangle \supset \langle a\rangle + \langle b\rangle$. $x \in \langle a\rangle + \langle b\rangle \implies x = a_x + b_x$ for $a_x \in \langle a\rangle, b_x \in \langle b\rangle$. But $a_x \in \langle a\rangle \implies a_x = r_a a$ for some $r_a \in R$, and $b_x \in \langle b\rangle \implies b_x = r_b b$ for some $r_b \in R$. So $x = r_a a + r_b b \iff x \in \langle a, b\rangle$.

We know that for $\langle a_1\rangle = \langle a_1\rangle$, and so by making this argument $n-1$ times we can build up by induction that $\langle a_1, \dots a_{n-1}, a_n\rangle = \langle a_1, \dots, a_{n-1}\rangle + \langle a_n\rangle$.

Exercise 5.47 (Shifrin 4.2.1)

(a) Prove that the function $\phi : \mathbb{Q}[\sqrt{2}] \to \mathbb{Q}[\sqrt{2}]$ defined by $\phi(a + b\sqrt{2}) = a - b\sqrt{2}$ is an isomorphism.
(b) Define $\phi : \mathbb{Q}[\sqrt{3}] \to \mathbb{Q}[\sqrt{7}]$ by $\phi(a + b\sqrt{3}) = a + b\sqrt{7}$. Is ϕ an isomorphism? Is there any isomorphism?

Solution. For (a), we first prove that it is a homomorphism.

$$\begin{aligned}
\phi((a+b\sqrt{2})+(c+d\sqrt{2})) &= \phi((a+c)+(b+d)\sqrt{2}) && (286)\\
&= (a+c)-(b+d)\sqrt{2} && (287)\\
&= (a-b\sqrt{2})+(c-d\sqrt{2}) && (288)\\
&= \phi(a+b\sqrt{2})+\phi(c+d\sqrt{2}) && (289)\\
\phi((a+b\sqrt{2})(c+d\sqrt{2})) &= \phi((ac+2bd)+(ad+bc)\sqrt{2}) && (290)\\
&= (ac+2bd)-(ad+bc)\sqrt{2} && (291)\\
&= (a-b\sqrt{2})(c-d\sqrt{2}) && (292)\\
&= \phi(a+b\sqrt{2})\times\phi(c+d\sqrt{2}) && (293)\\
\phi(1) &= 1 && (294)
\end{aligned}$$

This is injective since given that $a+b\sqrt{2}\neq c+d\sqrt{2}$, then at least $a\neq b$ or $c\neq d$, in which case $a-b\sqrt{2}\neq c-d\sqrt{2}$. Alternatively, we can see that the kernel is 0, so it must be injective. It is onto since given any $c+d\sqrt{2}$, the preimage is $c-d\sqrt{2}$. Therefore ϕ is an isomorphism.
For (b), no it is not an isomorphism since

$$\begin{aligned}
\phi((a+b\sqrt{3})(c+d\sqrt{3})) &= \phi((ac+3bd)+(ad+bc)\sqrt{3}) && (295)\\
&= (ac+3bd)+(ad+bc)\sqrt{7} && (296)\\
&\neq (ac+7bd)+(ad+bc)\sqrt{7} && (297)\\
&= (a+b\sqrt{7})(c+d\sqrt{7}) && (298)\\
&= \phi(a+b\sqrt{3})\phi(c+d\sqrt{3}) && (299)
\end{aligned}$$

We claim that there is no isomorphism. Assume that such ϕ exists. Then $\phi(1)=1$, and so $\phi(3)=\phi(1+1+1)=\phi(1)+\phi(1)+\phi(1)=1+1+1=3$. Now given $\sqrt{3}\in\mathbb{Q}[\sqrt{3}]$, we follows that

$$\phi(\sqrt{3})^2=\phi(3)=3 \tag{300}$$

and so $\phi(\sqrt{3})$ must map to the square root of 3 which must live in $\mathbb{Q}[\sqrt{7}]$. Assume such a number is $a+b\sqrt{7} \implies (a^2+7b^2)+(2ab)\sqrt{7}=\sqrt{3}$. This implies that $2ab=0$, leaving the rational term, but we know that $\sqrt{3}$ does not exist in the rationals, and so $\sqrt{3}$ does not exist.

Exercise 5.48 (Shifrin 4.2.3.a/c/e)

Establish the following isomorphisms (preferably, using Theorem 2.2):
(a) $\mathbb{R}[x]/\langle x^2+6\rangle\cong\mathbb{C}$
(b) $\mathbb{Z}_{18}/\langle\overline{6}\rangle\cong\mathbb{Z}_6$
(c) $\mathbb{Q}[x]/\langle x^2+x+1\rangle\cong\mathbb{Q}[\sqrt{3}i]$
(d) $\mathbb{Z}[x]/\langle 2x-3\rangle\cong\mathbb{Z}[\frac{3}{2}]=\{\frac{a}{b}\in\mathbb{Q}:b=2^j \text{ for some } j\geq 0\}\subset\mathbb{Q}$
(e) $F[x]/\langle x\rangle\cong F$
(f) $\mathbb{Z}_3\times\mathbb{Z}_4\cong\mathbb{Z}_{12}$

Solution. For all, we construct the ring homomorphism $\phi: R\to S$ with the appropriate kernel, and the result is immediate from the theorem.

a) Given $f\in\mathbb{R}[x]$ which is a Euclidean domain, we claim that the map $\phi_1: f(x)\mapsto r(x)$ where r is the remainder of f when divided by x^2+6, is a homomorphism. It is pretty easy to see that the map $\phi_2: f(x)=\sum_{k=0}^n a_kx^k\mapsto a_0+a_1i$ is also a homomorphism, and thus $\phi=\phi_2\circ\phi_1$ as the

composition of homomorphisms is also a ring homomorphism. ϕ_1 is a homomorphism since given $f, g \in \mathbb{R}[x]$, we can write them as $f(x) = d_1(x)(x^2+6) + r_1(x)$ and $g(x) = d_2(x)(x^2+6) + r_2(x)$. Therefore,

$$(f+g)(x) = f(x) + g(x) = (d_1(x) + d_2(x))(x^2+6) + (r_1 + r_2)(x) \tag{301}$$
$$(fg)(x) = f(x) \cdot g(x) = (d_1(x)(x^2+6) + r_1(x))(d_2(x)(x^2+6) + r_2(x)) \tag{302}$$
$$= (\dots)(x^2+6) + (r_1 + r_2)(x) \tag{303}$$
$$1 = 0(x^2+6) + 1 \tag{304}$$

Therefore ϕ is a homomorphism, and the kernel is simply all polynomials divisible by x^2+6, which is $\langle x^2+6 \rangle$.

c) We define $\phi(f) = f(\frac{-1+\sqrt{3}i}{2})$, where $\frac{-1+\sqrt{3}i}{2}$ is a root of x^2+x+1. Therefore, since $f \in \mathbb{R}$, $\frac{-1-\sqrt{3}i}{2}$ must also be a root and so the kernel is $\langle x^2+x+1 \rangle$. Second, we will show that it is a homomorphism.

$$\phi(f+g) = (f+g)\left(\frac{-1+\sqrt{3}i}{2}\right) = f\left(\frac{-1+\sqrt{3}i}{2}\right) + g\left(\frac{-1+\sqrt{3}i}{2}\right) = \phi(f) + \phi(g) \tag{305}$$
$$\phi(fg) = (fg)\left(\frac{-1+\sqrt{3}i}{2}\right) = f\left(\frac{-1+\sqrt{3}i}{2}\right) \cdot g\left(\frac{-1+\sqrt{3}i}{2}\right) = \phi(f) \cdot \phi(g) \tag{306}$$
$$\phi(1) = 1 \tag{307}$$

We are done.

e) Given $f(x) = \sum_{k=0}^{n} a_k x^k \in F[x]$, we show that $\phi : f \mapsto a_0$ is a homomorphism. Let f be as above and g have coeffients b_k from $k = 0 \dots m$.

$$\phi(f+g) = \phi\left(\sum_{k=0}^{\max\{n,m\}} (a_k + b_k) x^k \right) = a_0 + b_0 = \phi(f) + \phi(g) \tag{308}$$
$$\phi(fg) = \phi\left(\sum_{k=0}^{n+m} \left(\sum_{i=0}^{k} a_i b_{k-i} \right) x^k \right) = a_0 b_0 = \phi(f)\phi(g) \tag{309}$$
$$\phi(1) = 1 \tag{310}$$

So this is a homomorphism. Since $\langle x \rangle$ as all multiples of x consists of all polynomials with constant term $a_0 = 0$, we can see that $\ker(\phi) = 0$. Therefore we are done.

Exercise 5.49 (Shifrin 4.2.11.a/d)

True or false? (Give proofs or disproofs.)
(a) $\mathbb{Z}_2[x]/\langle x^2 \rangle \cong \mathbb{Z}_4$, or $\mathbb{Z}_2[x]/\langle x^2 \rangle \cong \mathbb{Z}_2 \times \mathbb{Z}_2$?
(b) Same questions for $\mathbb{Z}_2[x]/\langle x^2 + x \rangle$.
(c) Same questions for $\mathbb{Z}_2[x]/\langle x^2 + 1 \rangle$.
(d) $\mathbb{Z}_3[x]/\langle x^2 - 1 \rangle \cong \mathbb{Z}_3 \times \mathbb{Z}_3$?
(e) $\mathbb{Q}[x]/\langle x^2 - 1 \rangle \cong \mathbb{Q} \times \mathbb{Q}$?

Solution. Listed.

(a) False for both. The characteristic of $\mathbb{Z}_2[x]/\langle x \rangle$ is 2 since $1 + 1 = 0$, but the characteristic of $\mathbb{Z}_4$ is 4 since $1 + 1 + 1 + 1 = 0$, so false. As for $\mathbb{Z}_2 \times \mathbb{Z}_2$, note that $(0,1)$ and $(1,0)$ are zero divisors of each other where $(0,1) \cdot (1,0) = (0,0)$. However, the two zero divisors in $\mathbb{Z}_2[x]/\langle x \rangle$ are x and $x+1$, where $x^2 = (x+1)^2 = 0$. An isomomorphism $\phi : \mathbb{Z}_2[x]/\langle x \rangle \to \mathbb{Z}_2 \times \mathbb{Z}_2$ would have to

preserve $0 = \phi(0) = \phi(x^2) = \phi(x) \cdot \phi(x)$, but there are no nonzero elements $(a, b) \in \mathbb{Z}_2 \times \mathbb{Z}_2$ whose square is 0. Therefore, there cannot be an isomorphism.

(d) True. All elements of $\mathbb{Q}[x]/\langle x^2 - 1 \rangle$ are of form $a + bx$, with $a, b \in \mathbb{Q}$. We define the isomorphism $\phi(a + bx) = (a + b, a - b) \in \mathbb{Z}_3 \times \mathbb{Z}_3$. This is a homomorphism since

$$\begin{aligned}
\phi((a_1 + b_1x) + (a_2 + b_2x)) &= \phi((a_1 + a_2) + (b_1 + b_2)x) && (311)\\
&= (a_1 + a_2 + b_1 + b_2, a_1 + a_2 - b_1 - b_2) && (312)\\
&= (a_1 + b_1, a_1 - b_1) + (a_2 + b_2, a_2 - b_2) && (313)\\
&= \phi(a_1 + b_1x) + \phi(a_2 + b_2x) && (314)\\
\phi((a_1 + b_1x)(a_2 + b_2x)) &= \phi(a_1a_2 + (a_1b_2 + a_2b_1)x + b_1b_2x^2) && (315)\\
&= \phi((a_1a_2 + b_1b_2) + (a_1b_2 + a_2b_1)x) && (316)\\
&= (a_1a_2 + b_1b_2 + a_1b_2 + a_2b_1, a_1a_2 + b_1b_2 - a_1b_2 - a_2b_1) && (317)\\
&= (a_1 + b_1, a_1 - b_1)(a_2 + b_2, a_2 - b_2) && (318)\\
&= \phi(a_1 + b_1x)\phi(a_2 + b_2x) && (319)\\
\phi(1) &= 1 && (320)
\end{aligned}$$

This is also injective since given $a_1 + b_1x \neq a_2 + b_2x$, say that their images are the same. Then $a_1 + b_1 = a_2 + b_2$ and $a_1 - b_1 = a_2 - b_2$. Adding and subtracting the two equations, we have $2a_1 = 2a_2$ and $2b_1 = 2b_2$, which means the original elements were the same.

Exercise 5.50 (Shifrin 4.2.12)

Let R be a commutative ring, $I \subset R$ an ideal. Suppose $a \in R$, $a \notin I$, and $I + \langle a \rangle = R$ (see Exercise 4.1.17 for the notion of the sum of two ideals). Prove that $\bar{a} \in R/I$ is a unit.

Solution. Since $R = I + \langle a \rangle$, $1 \in R = I + \langle a \rangle$. So there exists $i \in I, ra \in \langle a \rangle$ s.t. $1 = i + ra \implies ra = 1 - i$. Therefore, in the quotient ring, $\bar{i} = 0$ and we have

$$\bar{r}\bar{a} = \bar{1} - \bar{0} = \bar{1} \tag{321}$$

and so $\bar{r}$ is a multiplicative inverse of $\bar{a}$. So $\bar{a}$ is a unit.

6 Domains

We can see that domains behave similarly to the integers, but with the missing property that $\times$ is commutative. This motivates the following definition of an integral domain, which can be seen as a generalization of the integers.

Definition 6.1 (Domain, Integral Domain)

A ring R with no zero divisors for every element is called a **domain**. An **integral domain** is a commutative domain R.[a]

[a] Almost always, we work with integral domains so we will default to this.

Example 6.1 (Domains vs Integral Domains)

We show some examples of domains and integral domains.

1. $(\mathbb{Z}, +, \times)$ is an integral domain
2. $(\mathbb{Q}, +, \times)$ is an integral domain.
3. $(\mathbb{R}, +, \times)$ is an integral domain.
4. Quaternions $\mathbb{H}$ are not commutative but are a domain.

Example 6.2 (Non-Domains)

Here are some examples of non-domains.

1. The ring of $n \times n$ matrices over any nonzero ring when $n \geq 2$ is not a domain. Given matrices A, B, if the image of B is in the kernel of A, then $AB = 0$.
2. The ring of continuous functions on the interval is not a domain. To see why, notice that given the piecewise functions

$$f(x) = \begin{cases} 1 - 2x & x \in [0, \frac{1}{2}] \\ 0 & x \in [\frac{1}{2}, 1] \end{cases}, \qquad g(x) = \begin{cases} 0 & x \in [0, \frac{1}{2}] \\ 2x - 1 & x \in [\frac{1}{2}, 1] \end{cases} \tag{322}$$

   $f, g \neq 0$, but $fg = gf = 0$.
3. A product of two nonzero commutative rings with unity $R \times S$ is not an integral domain since $(1, 0) \cdot (0, 1) = (0, 0) \in R \times S$.

The nice thing about not having zero divisors is that they give a nice cancellation property in rings. Therefore, a nonzero element that is not a zero divisor enjoys some of the properties of a unit without necessarily possessing a multiplicative inverse.

Lemma 6.1 (Cancellation Property of Elements)

In an integral domain R, if $ab = ac$, then either $a = 0$ or $b = c$.

Proof. We have

$$ab = ac \implies ab - ac = a(b - c) = 0 \tag{323}$$

and since R is a domain, one of the claims must follow.

This extends into ideals as well.

Theorem 6.2 (Cancellation Property of Ideals)

In an integral domain R and two ideals $I, J \subset R$, if $aI = aJ$, then either $a = 0$ or $I = J$.

Proof.

Another nice property is that the characteristic of an integral domain must be prime and that gcd's—while not yet unique—are now guaranteed to be associated.

Corollary 6.3 (Characteristic of Integral Domain)

If R is an integral domain, then $\text{char}(R)$ is either 0 or a prime number.

Proof. Let $m \in \mathbb{Z}$ be such that $\langle m \rangle = \ker f$. If $m = ab$, then $f(a)f(b) = f(m) = 0$. Since R is an integral domain, $f(a) = 0$ or $f(b) = 0$. Thus $d \in \ker f = \langle m \rangle$ or $e \in \ker f = \langle m \rangle \implies m$ is prime or 0.

However, the converse is not true, as there exist rings with prime characteristic that are not integral domains, e.g. $\mathbb{Z}_p \times \mathbb{Z}_p$.

Theorem 6.4 (Ideals of an Integral Domain)

R is an integral domain if and only if the only ideal $\langle 0 \rangle$ of R is a prime ideal.

Proof.

Theorem 6.5 (Primes are Irreducible)

In an integral domain R, a prime $p \in R$ is always irreducible.

Here is an alternative equivalent characterization of an integral domain.

Definition 6.2 (Regular Elements)

An element r of a ring R is **regular** if the mapping

$$\rho : R \longrightarrow R, \qquad x \mapsto xr \tag{324}$$

is injective for all $x \in R$.

Theorem 6.6 (Integral Domains w.r.t. Regularity)

An integral domain is a commutative ring where every element is regular.

6.1 Unique Factorization Domains

Definition 6.3 (Unique Factorization Domain)

A **unique factorization domain (UFD)** is an integral domain R in which every nonzero element $r \in R$ which is not a unit can be written as a finite product of irreducible elements $p_i \in R$ (not necessarily distinct).

$$r = p_1 p_2 \dots p_n \tag{325}$$

The decomposition is unique up to associates and permutations.

We have shown that in an integral domain, a prime is always irreducible. What about the converse?

Theorem 6.7 (Primes and Irreducibility are Equivalent in UFDs)

Given a UFD R, a nonzero element $p \in R$ is prime if and only if it is irreducible.

The next part is that GCD's are unique.[3] Furthermore, the following theorem gives us an algorithmic method of computing GCDs.

Theorem 6.8 (GCD in a UFD)

Let $a, b \in R$ be two nonzero elements of a UFD R and suppose

$$a = up_1^{e_1} p_2^{e_2} \dots p_n^{e_n}, \qquad b = vp_1^{f_1} p_2^{f_2} \dots p_n^{f_n} \tag{326}$$

are prime factorizations for a and b, where u, v are units, the primes $p_1, \dots, p_n$ are distinct, and the exponents $e_i, f_i \geq 0$. Then the element

$$d = p_1^{\min(e_1, f_1)} p_2^{\min(e_2, f_2)} \dots p_n^{\min(e_n, f_n)} \tag{327}$$

is the greatest common divisor of a and b.

Proof. Since the exponents of each of the primes occurring in d are no larger than the exponents occurring in the factorizations of both a and b, d divides a and b. To show that d is a greatest common divisor, let c be any common divisor of a and b, and we can factor it as

$$c = q_1^{g_1} q_m^{g_m} \dots q_m^{g_m} \tag{328}$$

Since each q_i divides c, it hence divides a and b, and we see that q_i must divide one of the primes p_j. In particular, up to associates the primes occurring in c must be a subset of the primes occurring in a and b.

$$\{q_1, \dots, q_m\} \subset \{p_1, \dots, p_n\} \tag{329}$$

Similarly, the exponents for the primes occurring in c must be no larger than those occurring in d. This implies that $c \mid d$.

[3] Though we can make a slightly more general statement about uniquness in *GCD domains*.

6.2 Principal Ideal Domains

Definition 6.4 (Principal Ideal Domain)

A **principal ideal domain**, also called a **PID**, is an integral domain in which every ideal is principal.

So a principal ideal domain is an integral domain by definition. It may seem that PIDs are an oddly specific structure to be studying separately, but this actually turns out to unlock a lot more nice properties that we are familiar with. The first is that GCDs are now unique, which is great. Second, we have Bezout's identity, saying that if x and y are elements of a PID without common divisors, then every element of the PID can be written in the form $ax + by$. Finally, and most importantly, any element of a PID has a unique decomposition into irreducible factors. We now introduce some examples of PIDs, which are not as trivial and should be introduced as theorems.

Example 6.3 (Integers and Polynomials over Fields are PIDs)

The following are all examples of principal ideal domains.

1. It is quite easy to see that any field $\mathbb{F}$ is a PID since the only two possible ideals are $\{0\}$ and $\mathbb{F}$, both of which are principal.
2. The ring of integers $\mathbb{Z}$ is a PID. If $I \subset \mathbb{Z}$ is an ideal, then if $I = \langle 0 \rangle$, then we're done. Otherwise, let $a \in I$ be the smallest positive integer in I. It is clear that $\langle a \rangle \subset I$. Now given an element $b \in I$, by the Euclidean algorithm we have $b = aq + r$ with $r < a$. Since $a, b \in I$, it follows that $r \in I$. But since $0 \leq r < a$ and a is the smallest positive integer, $r = 0$, and so $b = aq \implies b \in \langle a \rangle$.

Note that we have established the existence and uniqueness of the gcd in UFDs. In PIDs, we can say something stronger through *Bezout's lemma*.[4] It allows us to represent the gcd as a linear combination of ring elements.

Lemma 6.9 (Bezout's Lemma in PIDs)

Let R be a PID and $x, y \in R$. Then, there exists $a, b \in R$ s.t.

$$ax + by = d \tag{330}$$

Proof. We can prove this by showing that $\langle x \rangle + \langle y \rangle = \langle d \rangle$. We know that sums of ideals are ideals, and since R is a PID, $I = \langle x \rangle + \langle y \rangle$ must be principal, i.e. $I = \langle d \rangle$ for $d = a_0 x + b_0 y$ for some $a_0, b_0 \in R$. We claim that d is the gcd.

1. *d is a divisor.* We have $x \in I = \langle d \rangle$, which implies that $d \mid x$. Similarly for y, we have $d \mid y$.
2. *d is greatest.* If $c \mid x, c \mid y$, then $x = rc$ and $y = sc$, and substituting this in for d, we have

$$d = a_0 rc + b_0 sc \implies c \mid d \tag{331}$$

Note that in general, a UFD does not have to satisfy Bezout's lemma.

Example 6.4 (UFDs Doesn't Necessarily Satisfy Bezout's Lemma)

Take the UFD $\mathbb{Z}[x]$, with $2, x \in \mathbb{Z}[x]$. It clearly has gcd of 1, but there are no solutions to the equation

$$2f(x) + xg(x) = 1 \tag{332}$$

[4] Actually, this holds for a slightly more general structure called *Bezout rings*.

Corollary 6.10 (Ideals Generated by Irreducible Elements)

Let R be a PID and $I \subsetneq R$ be a proper ideal.
1. If $a \in I$ is irreducible, then $I = \langle a \rangle$.
2. a is irreducible iff $\langle a \rangle$ is maximal.

Proof. For the first claim, it is clear that $a \in I \implies \langle a \rangle \subset I$. Now we show that I cannot be strictly bigger. Assume that it was, i.e. take $b \in I \setminus \langle a \rangle$. Then, there exists $x, y \in R$ s.t.

$$ax + by = 1 \tag{333}$$

and so $1 \in I \implies I = R$, which contradicts the fact that I is proper. Therefore $I = \langle a \rangle$. This also proves the forward direction of the second claim. For the reverse direction, we prove the contrapositive by assuming that $\langle a \rangle$ is not maximal. Then there exists an ideal I s.t. $\langle a \rangle \subsetneq I \subsetneq R$. Since R is a PID, $I = \langle b \rangle$, and so $\langle a \rangle \subset \langle b \rangle \implies b \mid a$. Therefore, a is not irreducible.

We finally establish the hierarchy of PIDs.

Theorem 6.11 (PIDs are UFDs)

Every principal ideal domain is a unique factorization domain.

Proof. We show that it is impossible to find an infinite sequence $a_1, a_2, \dots$ s.t. a_i is divisible by a_{i+1} but is not an associate. Once done we can iteratively factor an element as we are guaranteed this process terminates. Suppose such a sequence exists. Then the a_i generate the sequence of distinct principal ideals $\langle a_1 \rangle \subset \langle a_2 \rangle \subset \dots$. Then union of these ideals is some principal ideal $\langle a \rangle$. So $a \in \langle a_n \rangle$ for some n by definition of containment in the intersection. But then for all $i \geq n$, this must mean that $\langle a_i \rangle = \langle a_n \rangle$, which implies that $\langle a \rangle$ can be obtained through a finite intersection, a contradiction.
Now we prove uniqueness. Each irreducible p generates a maximal deal $\langle p \rangle$ from 6.10. Next suppose an element of R has two factorizations.

$$p_1 p_2 \dots p_r = q_1 q_2 \dots q_s \tag{334}$$

Consider the ideals $\langle p_i \rangle, \langle q_i \rangle$. Relabel so that p_1 generates a minimal ideal amongst these (i.e. does not strictly contain another one of these ideals). Now we show that $\langle p_1 \rangle = \langle q_i \rangle$ for some i. Suppose not. Then $\langle p_1 \rangle$ does not contain any q_i, thus q_i is nonzero modulo $\langle p_1 \rangle$ for all i, which is a contradiction because the LHS of the above equation is zero modulo $\langle p_1 \rangle$.
Relabel so that $\langle p_1 \rangle = \langle q_1 \rangle$. Then $p_1 = w q_1$ for some unit u. Cancelling gives $u p_2 \dots p_r = q_2 \dots q_s$. The element $u p_2$ is also irreducible, so by induction we have that the factorization is unique.

Example 6.5 (UFD that is not a PID)

$\mathbb{Z}[x]$ is a UFD, but not a PID. We can see this in two ways.
1. The ideal $I = \langle 2, x \rangle \subset \mathbb{Z}[x]$ is not principal. Suppose it was. Then we have $\langle a \rangle = \langle 2, x \rangle$ for some $a \in \mathbb{Z}[x]$. So $a \mid 2$ and $a \mid x$. Note that $a \mid 2 \implies a = 1$ or $a = 2$, but $a = 2$ means that $a \nmid x$. So $a = 1$, and so $I = R$. However, $x + 1 \notin I$ since the multiples of x cannot affect the constant term, and multiples of 2 must be even. So $I \neq R$.
2. If it was a PID, then from 8.18 this would imply that $\mathbb{Z}$ is a field, which we know is not.

6.3 Euclidean Domains

We have seen that PIDs unlock a lot of familiar properties that we see in integers. In fact, pretty much everything holds except for the existence of Euclidean algorithm for factorization, which turns out to be extremely powerful. First, we define the notion of a *norm* on an integral domain R, which is no more than a measure of size. Note that it is different from the usual sense of norm in vector spaces, which satisfies the additional axioms of scalar multiplication and triangle inequality.

Definition 6.5 (Norm on an Integral Domain)

Given an integral domain R, a function $N : R \to \mathbb{N} \cup \{0\}$ with $N(0) = 0$ is called a **(Euclidean) norm** on the integral domain R.

Definition 6.6 (Euclidean Domain)

An integral domain R is a **Euclidean domain** if there is a norm N on R such that for any two elements $a, b \in R$ with $b \neq 0$, there exists elements $q, r \in R$, with

$$a = qb + r, \qquad r = 0 \text{ or } N(r) < N(b) \tag{335}$$

q is called the **quotient** and r is called the **remainder**.

The two prime examples are the integers and polynomials.

Example 6.6 (Integers)

$\mathbb{Z}$ is a Euclidean domain with Euclidean division, also called long division, defined

$$\begin{array}{r} 40 \\ 13\overline{)521} \\ \underline{52} \\ 01 \end{array}$$

Example 6.7 (Gaussian Integers)

The subring of $\mathbb{C}$, defined

$$\mathbb{Z}[i] \equiv \{a + bi \mid a, b \in \mathbb{Z}\} \tag{336}$$

is a Euclidean integral domain with respect to the norm

$$N(c) \equiv a^2 + b^2 \tag{337}$$

since $N(cd) = N(c)N(d)$ and the invertible elements of $\mathbb{Z}[i]$ are $\pm 1, \pm i$.

Example 6.8 (Dyadic Rationals)

The ring of rational numbers of the form $2^{-n}m,\ n \in \mathbb{Z}_+, m \in \mathbb{Z}$, is a Euclidean domain. To define the norm, we can first assume that m can be prime factorized into the form

$$m = \pm \prod_i p_i^{k_i},\ p \text{ prime} \tag{338}$$

and the norm is defined

$$N(\frac{m}{2^n}) \equiv 1 + \sum_i k_i \tag{339}$$

We must further show that division with remainder is possible, but we will not show it here.

The first implication of a division algorithm for an integral domain is that it forces every ideal of R to be principal.

Theorem 6.12 (Euclidean Domains are PIDs)

Let R be a Euclidean domain.
1. R is a principal ideal domain.
2. Every nonzero ideal $I \subset R$ is of the form $\langle d \rangle$, where $d \in I$ is an element of minimum norm.

Proof. Let I be an ideal. Then if I is the zero ideal, there is nothing to prove. Otherwise let d be a nonzero element I of minimum norm, which exists since the set $\{N(a) \mid a \in I\} \subset \mathbb{N}$ has minimum element by the Well Ordering Principl. Clearly $\langle d \rangle \subset I$. To show the reverse inclusion, let $a \in I$. Then we use the division algorithm to see that $a = qd + r$ with $r = 0$ or $N(r) < N(d)$. Then $r = a - qd$ and both $a, qd \in I$, which means $r \in I$. By the minimality of the norm of d, $r = 0$. Therefore $a = qd \in \langle d \rangle$, implying that $I \subset \langle d \rangle$.

Corollary 6.13 (Fundamental Theorem of Arithmetic)

$\mathbb{Z}$ is a unique factorization domain.

Proof. $\mathbb{Z}$ is a Euclidean domain, hence a PID, hence a UFD.

A useful fact that we will use later is to verify whether a quotient ring is a Euclidean domain. There is an analogous statement for fields.

Theorem 6.14 (Quotient Rings as Euclidean Domains)

Let R be a nontrivial commutative ring and $I \subset R$ an ideal. R/I is a Euclidean domain iff I is a prime ideal.

Proof. TBD

Finally we show a condition that a ring is not a Euclidean domain.

Definition 6.7 (Universal Side Divisor)

Given integral domain R, let $\tilde{R} = R^* \cup \{0\}$ be the set of units of R together with 0. An element $u \in R \setminus \tilde{R}$ is called a **universal side divisor** if for every $x \in R$ there is some $z \in \tilde{R}$ s.t. $u \mid x - z$ in R.

In other words, there is a type of division algorithm for u: Every x may be written $x = qu + z$, where z is either 0 or a unit. The existence of universal side divisors is a weakening of the Euclidean condition.

Theorem 6.15 (Euclidean Domains Contain Universal Side Divisors)

Let R be an integral domain that is not a field. If R is a Euclidean domain, then there are universal side divisors in R.

Proof. TBD

Example 6.9

We claim that the quadratic integer ring $R = \mathbb{Z}[\frac{1+\sqrt{19}}{2}]$ is not a Euclidean domain w.r.t. any norm.

6.4 Exercises

Exercise 6.1 (Dummit 7.1.11)

Prove that if R is an integral domain and $x^2 = 1$ for some $x \in R$ then $x = \pm 1$.

Solution. By subtracting 1 from both sides, we can factor to get

$$x^2 - 1 = (x+1)(x-1) = 0 \implies x = \pm 1 \tag{340}$$

where the last implication follows from neither being a zero divisor.

Exercise 6.2 (Dummit 7.1.15)

A ring R is called a *Boolean ring* if $a^2 = a$ for all $a \in R$. Prove that every Boolean ring is commutative.

Solution. Given $a, b \in R$, we have

$$ab = (ab)^2 = abab \tag{341}$$
$$ab = a^2b^2 \tag{342}$$

Now setting them equal to each other, we have $abab = a^2b^2 \implies ba = ab$.

Exercise 6.3 (Dummit 7.1.16)

Prove that the only Boolean ring that is an integral domain is $\mathbb{Z}/2\mathbb{Z}$.

Solution.

7 Fields

Our final structure is field, which seems to add only a few more conditions to a ring, but again unlocks more structure. Field theory is usually pretty tame compared to groups and rings. The purpose of this section is to really just introduce the definition of a field, plus the construction of the field of fractions—which will be useful for the construction of $\mathbb{Q}$ and the analysis of UFD polynomial rings—and finally ordered fields, which is nice again in the context of $\mathbb{Q}$ and $\mathbb{R}$.

Definition 7.1 (Field)

A **field** $(F, +, \times)$ is a commutative, associative ring where every nonzero element is a unit.

Lemma 7.1 (Properties of Addition)

The properties of addition hold in a field.
1. If $x + y = x + z$, then $y = z$.
2. If $x + y = x$, then $y = 0$.
3. If $x + y = 0$, then $y = -x$.
4. $(-(-x)) = x$.

Proof. For the first, we have

$$\begin{aligned} x + y = x + z &\implies -x + (x + y) = -x + (x + z) && \text{(addition is a function)} \\ &\implies (-x + x) + y = (-x + x) + z && (+ \text{ is associative}) \\ &\implies 0 + y = 0 + z && \text{(definition of additive inverse)} \\ &\implies y = z && \text{(definition of identity)} \end{aligned}$$

For the second, we can set $z = 0$ and apply the first property. For the third, we have

$$\begin{aligned} x + y = 0 &\implies -x + (x + y) = -x + 0 && \text{(addition is a function)} \\ &\implies (-x + x) + y = -x + 0 && (+ \text{ is associative}) \\ &\implies 0 + y = -x + 0 && \text{(definition of additive inverse)} \\ &\implies y = -x && \text{(definition of identity)} \end{aligned}$$

For the fourth, we simply follow that if y is an inverse of z, then z is an inverse of y. Therefore, $-x$ being an inverse of x implies that x is an inverse of $-x$. $-(-x)$ must also be an inverse of $-x$. Since inverses are unique[a], $x = -(-x)$.

[a] This is proved in algebra.

Lemma 7.2 (Properties of Multiplication)

The properties of multiplication hold in a field.
1. If $x \neq 0$ and $xy = xz$, then $y = z$.
2. If $x \neq 0$ and $xy = x$, then $y = 1$.
3. If $x \neq 0$ and $xy = 1$, then $y = x^{-1}$.
4. If $x \neq 0$, then $(x^{-1})^{-1} = x$.

Proof. The proof is almost identical to the first. Since $x \neq 0$, we can always assume that x^{-1} exists. For the first, we have

$$\begin{aligned} xy = xz &\implies x^{-1}(xy) = x^{-1}(xz) && \text{(multiplication is a function)} \\ &\implies (x^{-1}x)y = (x^{-1}x)z && (\times \text{ is associative}) \\ &\implies 1y = 1z && \text{(definition of multiplicative inverse)} \\ &\implies y = z && \text{(definition of identity)} \end{aligned}$$

For the second, we can set $z = 1$ and apply the first property. For the third, we have

$$\begin{aligned} xy = 1 &\implies x^{-1}(xy) = x^{-1}1 && \text{(multiplication is a function)} \\ &\implies (x^{-1}x)y = x^{-1}1 && (\times \text{ is associative}) \\ &\implies 1y = x^{-1}1 && \text{(definition of multiplicative inverse)} \\ &\implies y = x^{-1} && \text{(definition of identity)} \end{aligned}$$

For the fourth, we simply see that x^{-1} is a multiplicative inverse of both x and $(x^{-1})^{-1}$ in the group $(\mathbb{F} \setminus \{0\}, \times)$, and since inverses are unique, they must be equal.

Lemma 7.3 (Properties of Distribution)

For any $x, y, z \in \mathbb{F}$, the field axioms satisfy

1. $0 \cdot x = 0$.
2. If $x \neq 0$ and $y \neq 0$, then $xy \neq 0$.
3. $-1 \cdot x = -x$.
4. $(-x)y = -(xy) = x(-y)$.
5. $(-x)(-y) = xy$.

Proof. For the first, note that

$$0x = (0 + 0) \cdot x = 0x + 0x \tag{343}$$

and subtracting $0x$ from both sides gives $0 = 0x$. For the second, we can claim that $xy \neq 0$ equivalently claiming that it will have an identity. Since $x, y \neq 0$, their inverses exists, and we claim that $(xy)^{-1} = y^{-1}x^{-1}$ is an inverse. We can see that by associativity,

$$(y^{-1}x^{-1})(xy) = y^{-1}(x^{-1}x)y = y^{-1}y = 1 \tag{344}$$

For the third, we see that

$$0 = 0 \cdot x = (1 + (-1)) \cdot x = 1 \cdot x + (-1) \cdot x = x + (-1) \cdot x \tag{345}$$

which implies that $-1 \cdot x$ is the additive inverse. The fourth follows immediately from the third by the associative property. For the fifth we can see that

$$\begin{aligned} (-x)(-y) &= (-1)x(-1)y && \text{(property 3)} \\ &= (-1)(-1)xy && (\times \text{ is commutative}) \\ &= -1 \cdot (-xy) && \text{(property 3)} \\ &= -(-xy) && \text{(property 3)} \\ &= xy && \text{(addition property 4)} \end{aligned}$$

Theorem 7.4 (Fields are Euclidean Domains)

Every field is a Euclidean domain.

Proof. Given $x, y \in \mathbb{F}$, assume $xy = 0$ with $x \neq 0$. Since x is invertible,

$$0 = x^{-1}0 = x^{-1}(xy) = y \tag{346}$$

Now assuming that $y \neq 0$, since y is invertible,

$$0 = 0y^{-1} = (xy)y^{-1} = x \tag{347}$$

With this theorem, we have established the hierarchy in the beginning of this section. So as soon as we see a field, we can immediately apply everything we know, such as Euclidean division, unique factorization, GCDs, etc. The converse is not generally true except for finite fields.

Theorem 7.5 (Wedderburn's little theorem)

Every finite integral domain is a field.

Proof. Let R be a finite integral domain and $a \in R$ be nonzero. Since every element is regular, the map $x \mapsto ax$ is an injective function. Since R is finite this map is also surjective. In particular, there is some $b \in R$ s.t. $ab = 1$, i.e. a is a unit.

Theorem 7.6 (Integral Domains are Embedded in Fields)

An integral domain is a ring that is isomorphic to a subring of a field.

Proof. TBD

Theorem 7.7 (Ideals of Fields)

The only ideals that exist in a field $\mathbb{F}$ is $\{0\}$ and $\mathbb{F}$ itself.

Proof. Given a nonzero element $x \in \mathbb{F}$, every element of $\mathbb{F}$ can be expressed in the form of ax or xa for some $a \in \mathbb{F}$.

The ring $\mathbb{Z}_n$ has all the properties of a field except the property of having inverses for all of its nonzero elements. This leads to the following theorem.

Theorem 7.8 (Quotient Rings as Fields)

Let R be a nontrivial commutative ring and $I \subset R$ an ideal. R/I is a field iff I is a maximal ideal.

Proof. TBD

Corollary 7.9 (Integer Quotient Rings as Finite Fields)

The ring $(\mathbb{Z}_n, +, \times)$ is a field if and only if n is a prime number.

Proof. This proof is a one-liner given the previous theorem, but let's provide an alternative proof.

1. $(\rightarrow)$ We prove the contrapositive. Assume that n is composite $\implies$ $n = kl$ for $k, n \in \mathbb{N}$ $\implies$ $k, n \neq 0$, but
$$[k]_n[l]_n = [kl]_n = [n]_n = 0 \tag{348}$$
meaning that $\mathbb{Z}_n$ contains 0 divisors and is not a field.
2. $(\leftarrow)$ Given that n is prime, let $[a]_n \neq 0$, i.e. $[a]_n \neq [0]_n, [1]_n$. The set of n elements
$$[0]_n, [a]_n, [2a]_n, ..., [(n-1)a]_n \tag{349}$$
are all distinct. Indeed, if $[ka]_n = [la]_n$, then $[(k-l)a]_n = 0$ $\implies$ $n = (k-l)a$ $\iff$ n is not prime. Since the elements are distinct, exactly one of them must be $[1]_n$, say $[pa]_n$ $\implies$ the inverse $[p]_n$ exists.

Corollary 7.10 (Invertibility in $\mathbb{Z}_n$)

For any n, $[k]_n$ is invertible in the ring $\mathbb{Z}_n$ if and only if n and k are relatively prime.

Proof. TBD

We will talk about finite fields again, which are extremely important in Galois theory and in practical applications in e.g. cryptography.

7.1 Field of Fractions and the Rationals

Given an integral domain, there is a common way to construct a field from it. We simply just "add" all the multiplicative inverses. Doing so with the integers and polynomials creates the field of rational numbers and rational functions. To formalize this a bit more, we claim that any integral domain R is a subring of a larger field F. Actually, we can prove an even stronger claim about *commutative rings*, i.e. every commutative ring is a subring of a larger ring S in which every nonzero element of R that is not a zero divisor is a unit in S.

Definition 7.2 (Ring/Field of Fractions)

Let R be a commutative ring and D be any nonempty subset of R^a that

1. does not contain 0,
2. does not contain zero divisors,
3. is closed under multiplication.

Then, there exists a commutative subring $Q \supset R$, called the **ring of fractions**, with the properties.

1. *Fraction.* Every element of Q is of the form rd^{-1}, for some $r \in R, d \in D$. If $D = R \setminus \{0\}$, then Q is a field, called the **field of fractions**.
2. *Uniqueness.* Q is the smallest ring containing R in which every elements of D become units in the following sense. Let S be any commutaticve ring with identity and let $\varphi : R \to S$ be any injective ring homomorphism such that $\varphi(d)$ is a unit in S for every $d \in D$. Then there is an injective homomorphism $\phi : Q \to S$ s.t. $\phi|_R = \varphi$.[b]

[a] This is basically a subgroup of units.

[b] In other words, any ring containing an isomorphic copy of R in which all the elements of D become units must also contain an isomorphic copy of Q.

Proof. Let $F = \{(r, d) \in R \times D \mid r \in R, d \in D\}$ and define the relation $\sim$ on F by

$$(r, d) \sim (s, e) \iff re = sd \tag{350}$$

It is indeed an equivalence relation.

1. *Reflexive.* $(r, d) \sim (r, d)$ since $rd = rd$.
2. *Symmetric.* Let $(r, d) \sim (s, e)$. Then $re = sd \iff sd = re$, and so $(s, e) \sim (r, d)$.
3. *Transitive.* Let $(r, d) \sim (s, e)$ and $(s, e) \sim (t, f)$. Then $re - sd = 0$ and $sf - te = 0$. Multiplying the first and second equations by f and d respectively and adding them gives $(rf - td)e = 0$. $e \neq 0$ or e is not a zero divisor, so $rf = td \iff (r, d) \sim (t, f)$.

Let us denote the equivalence class of (r, d) as $\frac{r}{d}$, and let Q be the set of equivalence classes under $\sim$. Note that $\frac{r}{d} = \frac{re}{de}$ for all $e \in D$, since D is closed under multiplication.[a] We now define addition and multiplication on Q as

$$\frac{a}{b} + \frac{c}{d} \coloneqq \frac{ad + bc}{bd}, \qquad \frac{a}{b} \times \frac{c}{d} = \frac{ac}{bd} \tag{351}$$

which is again well defined since D is closed under multiplication and do not depend on the choice of representatives from the equivalence class.

1. Verify the additive identity.

$$(a, b) + (0, c) = (ac + 0b, bc) = (ac, bc) \sim (a, b) \tag{352}$$

2. Verify the multiplicative identity.

$$(a, b) \times (c, c) = (ac, bc) \sim (a, b) \tag{353}$$

3. Additive inverse is actually an inverse.

$$(a, b) + (-a, b) = (ab + (-ba), bb) = (0, bb) \sim (0, 1) \tag{354}$$

4. Multiplicative inverse is actually an inverse.

$$(a, b) \times (b, a) = (ab, ba) = (ab, ab) \sim (1, 1) \tag{355}$$

5. Addition is commutative.

$$(a, b) + (c, d) = (ad + bc, bd) = (cb + ad, bd) = (c, d) + (a, b) \tag{356}$$

6. Addition is associative.

$$\begin{aligned} (a, b) + ((c, d) + (e, f)) &= (a, b) + (cf + de, df) && (357)\\ &= (adf + bcf + bde, bdf) && (358)\\ &= (ad + bc, bd) + (e, f) && (359)\\ &= ((a, b) + (c, d)) + (e, f) && (360) \end{aligned}$$

7. Multiplication is commutative.

$$(a, b) \times (c, d) = (ac, bd) = (ca, db) = (c, d) \times (a, b) \tag{361}$$

8. Multiplication is associative.

$$\begin{aligned} (a, b) \times ((c, d) \times (e, f)) &= (a, b) \times (ce, df) && (362)\\ &= (ace, bdf) && (363)\\ &= (ac, bd) \times (e, f) && (364)\\ &= ((a, b) \times (c, d)) \times (e, f) && (365) \end{aligned}$$

9. Multiplication distributes over addition.

$$\begin{aligned}(a,b)\times((c,d)+(e,f)) &= (a,b)\times(c,d)+(a,b)\times(e,f) && (366)\\ &= (ac,bd)+(ae,bf) && (367)\\ &= (abcf+abde,b^2df) && (368)\\ &= (acf+ade,bdf) && (369)\\ &= (a,b)\times(cf+de,df) && (370)\end{aligned}$$

Next we embed R into Q by defining

$$\iota : R \to Q, \qquad \iota(r) = \frac{rd}{d} \text{ for any } d \in D \tag{371}$$

Since $\frac{rd}{d} = \frac{re}{e}$ for all $d, e \in D$, $\iota(r)$ does not depend on the choice of $d \in D$ (for now, we assume that $1 \in D$ and continue the proof with this assumption, though the general case is also pretty simple). Since D is closed under multiplication, one checks directly that ι is a ring homomorphism.

1. Preservation of addition.

$$\begin{aligned}\iota(a)+_Q\iota(b) &= (a,1)+_Q(b,1) && (372)\\ &= (1a+_R1b,1^2) && (373)\\ &= (a+_Rb,1) && (374)\\ &= \iota(a+_Rb) && (375)\end{aligned}$$

2. Preservation of multiplication.

$$\begin{aligned}\iota(a)\times_Q\iota(b) &= (a,1)\times_Q(b,1) && (376)\\ &= (a\times_Rb,1^2) && (377)\\ &= (a\times_Rb,1) && (378)\\ &= \iota(a\times_Rb,1) && (379)\end{aligned}$$

3. Preservation of multiplicative identity.

$$\iota(1_R) = (1,1) = 1_Q \tag{380}$$

We then prove ι is injective since

$$\iota(r) = 0 \iff \frac{rd}{d} = \frac{0}{d} \iff rd^2 = 0 \iff r = 0 \tag{381}$$

because d—and hence d^2— is neither 0 nor a zero divisor. We therefore have $\iota(R) \simeq Q$.

Next, note that each $d \in D$ has a multiplicative inverse in Q. That is, if d is represented by the fraction $\frac{de}{e}$, then its multiplicative inverse is $\frac{e}{de}$. Therefore, every element of Q of the form rd^{-1} for some $r \in R, d \in D$.[b] It follows immediately that if $D = R \setminus \{0\}$, then every nonzero element of Q is a unit and so Q is a field.

For the uniqueness property, assume that $\varphi : R \to S$ is an injective ring homomorphism such that $\varphi(d)$ is a unit in S for all $d \in D$. Extend φ to a map $\phi : Q \to S$ by defining

$$\phi(rd^{-1}) = \varphi(r)\varphi(d)^{-1} \tag{382}$$

for all $r \in R, d \in D$. This map is well defined, since $rd^{-1} = se^{-1}$ implies $re = sd \implies \varphi(r)\varphi(e) = \varphi(s)\varphi(d)$, and so

$$\phi(rd^{-1}) = \varphi(r)\varphi(d)^{-1} = \varphi(s)\varphi(e)^{-1} = \phi(se^{-1}) \tag{383}$$

ϕ is indeed a ring homomorphism, and it is injectivce since $rd^{-1} \in \ker\phi$ implies $r \in \ker\phi \cap R = \ker\varphi$. Since φ is injective this forces r and hence rd^{-1} to be 0.

[a] This is why we can't multiply by $\frac{0}{0}$.

[b] Since $(r,d) = (r, \frac{de}{e}) = (r \cdot \frac{e}{de}, 1) = (rd^{-1}, 1) \in Q$.

Therefore, non-zero divisors get upgraded to units, while zero-divisors... well stay the same. The relative properties of these are quite simple, since R being an integral domain implies that it will have a *field* of fractions. We now derive the rational numbers as a field of fractions of the integers, which is straightforward.

Definition 7.3 (Rational Numbers)

The field of fractions of $\mathbb{Z}$ is called the **rational numbers**. More specifically, the **rational numbers** $(\mathbb{Q}, +_{\mathbb{Q}}, \times_{\mathbb{Q}})$ is the quotient space on $\mathbb{Z} \times \mathbb{Z} \setminus \{0\}$ with the equivalence relation $\sim$

$$(a,b) \sim (c,d) \iff a \times_{\mathbb{Z}} d = b \times_{\mathbb{Z}} c \tag{384}$$

and the operation defined

1. The additive and multiplicative identities are

$$0_{\mathbb{Q}} \coloneqq (0_{\mathbb{Z}}, a), \quad 1_{\mathbb{Q}} \coloneqq (a,a) \tag{385}$$

2. Addition on $\mathbb{Q}$ is defined

$$(a,b) +_{\mathbb{Q}} (c,d) \coloneqq \big((a \times_{\mathbb{Z}} d) +_{\mathbb{Z}} (b \times_{\mathbb{Z}} c), b \times_{\mathbb{Z}} d\big) \tag{386}$$

3. The additive inverse is defined

$$-(a,b) \coloneqq (-a,b) \tag{387}$$

4. Multiplication on $\mathbb{Q}$ is defined

$$(a,b) \times_{\mathbb{Q}} (c,d) \coloneqq \big(a \times_{\mathbb{Z}} c, b \times_{\mathbb{Z}} d\big) \tag{388}$$

5. The multiplicative inverse is defined

$$(a,b)^{-1} \coloneqq (b,a) \tag{389}$$

So for every commutative ring there is an associated ring of fractions. A natural question to ask is whether this is unique. Apparently, it is not.

Corollary 7.11

Let R be an integral domain and let Q be the field of fractiohns of R. If a field F contains a subring R' isomorphic to R, then the subfield of F generated by R' is isomorphic to Q.

Proof.

Lemma 7.12 (Rationals are a Minimal Field)

Every subfield of $\mathbb{C}$ contains $\mathbb{Q}$.

Proof. Must contain 0 and 1. Keep adding 1 and inverting it to get $\mathbb{Z}$. Now $\mathbb{Z}$ must contain units so $1/n$ also contained. Then multiply the elements to get $\mathbb{Q}$.

Once we define polynomials in the next section, we will construct the field of rational functions in the same manner.

7.2 Ordered Fields

Great, so we have established that $\mathbb{Q}$ is a field. The next property we want to formalize is order. There are countless ways to do it, but I just take the difference and claim that it is greater than 0. Note that given a set, we can really put whatever order we want on it. However, consider the field with the following order.

$$\mathbb{F} = \{0, 1\},\ 0 < 1 \tag{390}$$

This does not behave well with respect to its operations because for example if we have $0 < 1$, then adding the same element to both sides should preserve the ordering. But this is not the case since $0+1 = 1 > 1+1 = 0$. While it may be easy to define an order, we would like it to be an ordered field.

Definition 7.4 (Ordered Field)

An **ordered field** is a field that has an order satisfying
1. $y < z \implies x + y < x + z$ for all $x \in \mathbb{F}$.
2. $x > 0, y > 0 \implies xy > 0$.

Theorem 7.13 (Properties of an Ordered Field)

In an totally ordered field,
1. $x > 0 \implies -x < 0$.
2. $x \neq 0 \implies x^2 > 0$.
3. If $x > 0$, then $y < z \implies xy < xz$.

Proof. The first property is a single-liner

$$0 < x \implies 0 + -x < x + -x \implies -x < 0 \tag{391}$$

For the second property, it must be the case that $x > 0$ or $x < 0$. If $x > 0$, then by definition $x^2 > 0$. If $x < 0$, then

$$x^2 = 1 \cdot x^2 = (-1)^2 \cdot x^2 = (-1 \cdot x)^2 = (-x)^2 \tag{392}$$

and since $-x > 0$ from the first property, we have $x^2 = (-x)^2 > 0$. For the third, we use the distributive property.

$$\begin{aligned} y < z &\implies 0 < z - y && (393)\\ &\implies 0 = x0 < x(z - y) = xz - xy && (394)\\ &\implies xy < xz && (395) \end{aligned}$$

Theorem 7.14 (Ordered Field Structure)

Second, $\mathbb{Q}$ is an ordered field. The order $\leq_{\mathbb{Q}}$ defined on the rationals as

$$(a, b) \leq_{\mathbb{Q}} (c, d) \iff ad \leq_{\mathbb{Z}} bc \tag{396}$$

is a total order. Remember that WLOG we can assume $b, d > 0$.

Proof. For the order property, we have
1. Reflexive.
$$(a, b) \leq_{\mathbb{Q}} (a, b) \iff ab \leq_{\mathbb{Z}} ab \tag{397}$$
2. Antisymmetric.
$$(a, b) \leq_{\mathbb{Q}} (c, d) \implies ad \leq_{\mathbb{Z}} bc (c, d) \leq_{\mathbb{Q}} (a, b) \implies bc \leq_{\mathbb{Z}} ad \tag{398}$$

This implies that both $ad = bc$, which by definition means that they are in the same equivalence class.

3. Transitivity. Assume that $(a, b) \leq (c, d)$ and $(c, d) \leq (e, f)$. Then, we notice that $b, d, f > 0$ and therefore by the ordered ring property[a] of $\mathbb{Z}$, we have

$$(a, b) \leq_{\mathbb{Q}} (c, d) \implies ad \leq_{\mathbb{Z}} bc \implies adf \leq_{\mathbb{Z}} bcf \tag{399}$$
$$(c, d) \leq_{\mathbb{Q}} (e, f) \implies cf \leq_{\mathbb{Z}} de \implies bcf \leq_{\mathbb{Z}} bde \tag{400}$$

Therefore from transitivity of the ordering on $\mathbb{Z}$ we have $adf \leq bde$. By the ordered ring property[b] we have $0 \leq bde - adf = d(be - af)$. But notice that $d > 0$ from our definition of rationals, and therefore it must be the case that $0 \leq be - af \implies af \leq_{\mathbb{Z}} be$, which by definition means $(a, b) \leq_{\mathbb{Q}} (e, f)$.

For the ordered field property, we have

1. Assume that $y = (a, b) \leq (c, d) = z$. Let $x = (e, f)$. Then $x+y = (af+be, bf)$, $x+z = (cf+de, df)$. Therefore

$$(af + be)df = adf^2 + bedf \tag{401}$$
$$\leq bcf^2 + bedf \tag{402}$$
$$= (cf + de)bf \tag{403}$$

But $(af+be)df = (cf+de)bf$ is equivalent to saying $(af+be, bf) \leq_{\mathbb{Q}} (cf+de, df)$, i.e. $x+y \leq x+z$!

2. Let $x = (a, b), y = (c, d)$. Since $0 < x, 0 < y$, by construction this means that $0 < a, 0 < c$ (since $b, d > 0$ in the canonical rational form). By the ordered ring property of the integers, $0 < ac$. So

$$0 < ac \iff 0 \cdot bd < ac \cdot 1 \iff (0, 1) < (ac, bd) \iff 0_{\mathbb{Q}} < (a, c) \times_{\mathbb{Q}} (b, d) = xy \tag{404}$$

[a] If $a \leq b$ and $0 \leq c$, then $ac \leq bc$.
[b] If $a \leq b$, then $a + c \leq b + c$.

Great, so we have shown that the rationals have an ordered field structure and that there is a canonical ring embedding from $\mathbb{Z}$ to $\mathbb{Q}$. It remains to show that this ring homomorphism is an *ordered* ring homomorphism.

Theorem 7.15 (Canonical Injection of $\mathbb{Z}$ to $\mathbb{Q}$ is an Ordered Ring Homomorphism)

The canonical injection $\iota : \mathbb{Z} \to \mathbb{Q}$ defined $\iota(a) = (a, 1)$[a] is an ordered ring homomorphism. That is, for $a, b \in \mathbb{Z}$,

$$a \leq_{\mathbb{Z}} b \iff \iota(a) \leq_{\mathbb{Q}} \iota(b) \tag{405}$$

[a] which is well defined since we can arbitrarily choose the denominator.

Proof. We have already proved that this is a ring homomorphism. To show that it preserves the order, we have

$$a \leq_{\mathbb{Z}} b \iff a \cdot 1 \leq_{\mathbb{Z}} b \cdot 1 \tag{406}$$
$$\iff (a, 1) \leq_{\mathbb{Q}} (b, 1) \tag{407}$$
$$\iff \iota(a) \leq_{\mathbb{Q}} \iota(b) \tag{408}$$

Theorem 7.16 (Finite Fields)

There are no finite ordered fields.

Proof. Assume $\mathbb{F}$ is such an ordered field. It must be the case that $0, 1 \in \mathbb{F}$, with $0 < 1$. Therefore, we also have $0 + 1 < 1 + 1 \implies 1 < 1 + 1$. Repeating this we get

$$0 < 1 < 1 + 1 < 1 + 1 + 1 < \dots \tag{409}$$

where these elements must be distinct (since only one of $>, <, =$ must be true for a totally ordered set). Since this can be done for a countably infinite number of times, $\mathbb{F}$ cannot be finite.

Great, so we have pretty much constructed the rational numbers, with the exception that we still need the topology/metric/norm on these numbers, but this won't be too relevant for now.

7.3 The Real Numbers

The next step is to formally construct the real numbers. There are generally two ways of doing this: with Dedekind cuts, with Cauchy sequences, or with compact nested intervals. Using Cauchy sequences or compactness requires us to introduce the metric and the topology, while Dedekind cuts is purely based on the order which we have established. To make the construction as minimal as possible, I will use the Dedekind cuts method, and in topology/analysis, we can compare all three of these methods (and determine their equivalence!).

Definition 7.5 (Dedekind Cut)

A **Dedekind cut** is a partition of the rationals $\mathbb{Q} = A \sqcup A'$ satisfying the three properties.[a]

1. $A \neq \emptyset$ and $A \neq \mathbb{Q}$.[b]
2. $x < y$ for all $x \in A, y \in A'$.
3. The maximum element of A does not exist in $\mathbb{Q}$.

The minimum of A' may exist in $\mathbb{Q}$, and if it does, the cut is said to be **generated** by $\min A'$.

[a] This can really be defined for any totally ordered set.
[b] By relaxing this property, we can actually complete $\mathbb{Q}$ to the extended real number line.

Note that in $\mathbb{Q}$, there will be two types of cuts:

1. ones that are generated by rational numbers, such as

$$A = \{x \in \mathbb{Q} \mid x < 2/3\}, A' = \{x \in \mathbb{Q} \mid x \geq 2/3\} \tag{410}$$

2. and the ones that are not

$$A = \{x \in \mathbb{Q} \mid x^2 < 2\}, A' = \{x \in \mathbb{Q} \mid x^2 \geq 2\} \tag{411}$$

We can intuitively see that the set of all Dedekind cuts (A, A') will "extend" the rationals into a bigger set. We can then define some operations and an order to construct this into an ordered field, and finally it will have the property that we call "completeness."

Definition 7.6 (Dedekind Completeness)

A totally ordered algebraic field $\mathbb{F}$ is **complete** if every Dedekind cut of $\mathbb{F}$ is generated by an element of $\mathbb{F}$.

Theorem 7.17 (Rationals are Not Dedekind-Complete)

$\mathbb{Q}$ is not Dedekind-complete.

Proof. The counter-example is given above for the cut

$$A = \{x \in \mathbb{Q} \mid x^2 < 2\}, A' = \{x \in \mathbb{Q} \mid x^2 \geq 2\} \tag{412}$$

Now we have the tools to define the reals, giving us the beefy theorem.

Theorem 7.18 (Reals as the Dedekind-Completion of Rationals)

Let $\mathbb{R}$ be the set of all Dedekind cuts (A, A') of $\mathbb{Q}$ of $\mathbb{Q}$. For convenience we can uniquely represent (A, A') with just A since $A' = \mathbb{Q} \setminus A$. By doing this we can intuitively think of a real number as being represented by the set of all smaller rational numbers. Let A, B be two Dedekind cuts. Then, we define the following order and operations.

1. *Order.* $A \leq_{\mathbb{R}} B \iff A \subset B$.
2. *Addition.* $A +_{\mathbb{R}} B \coloneqq \{a +_{\mathbb{Q}} b \mid a \in A, b \in B\}$.
3. *Additive Identity.* $0_{\mathbb{R}} \coloneqq \{x \in \mathbb{Q} \mid x < 0\}$.
4. *Additive Inverse.* $-B \coloneqq \{a - b \mid a < 0, b \in (\mathbb{Q} \setminus B)\}$.
5. *Multiplication.* If $A, B \geq 0$, then we define $A \times_{\mathbb{R}} B \coloneqq \{a \times_{\mathbb{Q}} b \mid a \in A, b \in B, a, b \geq 0\} \cup 0_{\mathbb{R}}$. If A or B is negative, then we use the identity $A \times B = -(A \times_{\mathbb{R}} -B) = -(-A \times_{\mathbb{R}} B) = (-A \times_{\mathbb{R}} -B)$ to convert A, B to both positives and apply the previous definition.
6. *Multiplicative Identity.* $1_{\mathbb{R}} = \{x \in \mathbb{Q} \mid x < 1\}$.
7. *Multiplicative Inverse.* If $B > 0$, $B^{-1} \coloneqq \{a \times_{\mathbb{Q}} b^{-1} \mid a \in 1_{\mathbb{R}}, b \in (\mathbb{Q} \setminus B)\}$. If B is negative, then we compute $B^{-1} = -((-B)^{-1})$ by first converting to a positive number and then applying the definition above.

We claim that $(\mathbb{R}, +_{\mathbb{R}}, \times_{\mathbb{R}}, \leq_{\mathbb{R}})$ is a totally ordered field, and the canonical injection $\iota : \mathbb{Q} \to \mathbb{R}$ defined

$$\iota(q) = \{x \in \mathbb{Q} \mid x < q\} \tag{413}$$

is an ordered field isomorphism. Finally, by construction $\mathbb{R}$ is Dedekind-complete.

Definition 7.7 (Least Upper Bound Property)

A totally ordered set $\mathbb{F}$ has the **least upper bound** property if every nonempty set of F having an upper bound must have a least upper bound (supremum) in F.

Theorem 7.19 (Dedekind Completeness Equals Least-Upper-Bound Property)

Dedekind completeness is equivalent to the least upper bound property.

Proof.

It is also important to note that $\mathbb{R}$ satisfies the Archimidean principle, which is fundamental in analysis, and that Cauchy/nested interval completeness does *not* imply Archimidean, while Dedekind-completeness does. However, this again is not very relevant in a purely algeraic sense.

Definition 7.8 (Archimidean Principle)

An ordered ring $(X, +, \cdot, \leq)$ that embeds the naturals $\mathbb{N}$[a] is said to obey the **Archimedean principle** if given any $x, y \in X$ s.t. $x, y > 0$, there exists an $n \in \mathbb{N}$ s.t. $\iota(n) \cdot x > y$. Usually, we don't care about the canonical injection and write $nx > y$.

[a] as in, there exists an ordered ring homomorphism $\iota : \mathbb{N} \to X$

By the canonical injections $\mathbb{N} \to \mathbb{Z} \to \mathbb{Q} \to \mathbb{R}$, we can talk about whether this set has the Archimedean property. In fact Dedekind completeness does imply it.

Theorem 7.20 (Reals are Archimidean)

$\mathbb{R}$ satisfies the Archimedean principle.

Proof. Assume that this property doesn't hold. Then for any fixed x, $nx < y$ for all $n \in \mathbb{N}$. Consider the set

$$A = \bigcup_{n \in \mathbb{N}} (-\infty, nx), \quad B = \mathbb{R} \setminus A \tag{414}$$

A by definition is nonempty, and B is nonempty since it contains y. Then, we can show that $a \in A, b \in B \implies a < b$ using proof by contradiction. Assume that there exists $a' \in A, b' \in B$ s.t. $a' > b'$. Since $a' \in A$, there exists a $n' \in \mathbb{N}$ s.t. $a' \in (-\infty, n'x) \iff a' < n'x$. But by transitivity of order, this means $b' < n'x \iff b' \in (-\infty, n'x) \implies b' \in A$.
Going back to the main proof, we see that A is upper bounded by y, and so by the least upper bound property it has a supremum $z = \sup A$.

1. If $z \in A$, then by the induction principle[a] $z + x \in A$, contradicting that z is an upper bound.
2. If $z \notin A$, then by the induction principle[b] $z - x \notin A \implies z - x \in B$. Since every element of B upper bounds A and since $x > 0$, this means that $z - x < z$ is a smaller upper bound of A, contradicting that z is a least upper bound.

Therefore, it must be the case that $nx > y$ for some $n \in \mathbb{N}$.

[a] Note that $\mathbb{N}$ is defined recursively as $1 \in \mathbb{N}$ and if $n \in \mathbb{N}$, then $n + 1 \in \mathbb{N}$.
[b] The contrapositive of the recursive definition of $\mathbb{N}$ is: if $n \notin \mathbb{N}$, then $n - 1 \notin \mathbb{N}$.

7.4 Exercises

8 Polynomial Rings

In ring theory, the idea of *adjoining* an existing ring R with an arbitrary element x will appear frequently. That is, given a ring R and some x, can we try and construct a new ring S, denoted $R[x]$ that is the minimal ring containing both R and x? What kind of elements would be in $R[x]$?

1. Since $R \subset R[x]$, it must be the case that for $a \in R$, $a \in R[x]$.
2. Since $x \in R[x]$, it must be the case that $ax \in R[x]$ for all $a \in R$.
3. $x \times x = x^2 \in R[x]$, so it must be the case that $ax^2 \in R[x]$
4. In general, for any $n \in \mathbb{N}$, $x^n \in R[x]$, so it must be the case that $ax^n \in R[x]$.

If we add these terms up, we have elements of the general form

$$a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0 \tag{415}$$

This is what we refer to as a polynomial, and it indeed does have a ring structure. More generally, given a set of elements $S = \{x_1, \ldots, x_n\}$, $R[S]$ can be defined accordingly. Now let's formally define them.

Definition 8.1 (Polynomial Ring)

Given ring R and a set of **indeterminates** $S = \{x_1, \ldots, x_n\}$, the **ring of polynomials** $R[S] = R[x_1, \ldots, x_n]$ is defined in the equivalent ways.

1. It is the minimal ring containing R as a subring and S.
2. It is the ring of formal expressions of the form

$$f(x_1, \ldots, x_n) = \sum_{0 \leq k_i \leq n} a_{k_1 \ldots k_n} x_1^{k_1} x_2^{k_2} \ldots x_n^{k_n}, \quad a_i \in R \tag{416}$$

which for univariate polynomials simplifies to

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0, \quad a_i \in R \tag{417}$$

From both definitions is becomes clear through the ring properties that

1. Addition is defined component-wise: $ax^i + bx^i = (a+b)x^i$.
2. Multiplication is defined component-wise: $x^i x^j = x^{i+j}$.
3. Additive and multiplicative identities is 0 and 1.

Note that x is just a formal symbol, whose powers x^i are just placeholders for the corresponding coefficients a_i so that the given formal expression is a way to encode the finitary sequence. $(a_0, a_1, a_2, ..., a_n)$. Two polynomials are equal if and only if the sequences of their corresponding coefficients are equal. Also note that unlike a function, which we write as f, for a polynomial we should write the indeterminate $f(x)$. We can however interpret $f(x)$ as a function as well (which may not be unique), and doing so allows us to determine special properties of $f(x) \in R[x]$. Before we move on, let's get some terms out of the way.

Definition 8.2 (Some Terms for Polynomials)

Given a univariate polynomial $f(x) \in R[x]$.

1. The **leading coefficient** is the last nonzero coefficient
2. The **degree** of f—denoted $\deg f$—is the index of the leading coefficient.
3. A **monomial** is a polynomial of a single term $a_j x^j$.
4. A **linear** polynomial is a polynomial of degree 1.
5. A **quadratic** polynomial is a polynomial of degree 2.
6. A **cubic** polynomial is a polynomial of degree 3.

We need to be very careful about the properties that hold for polynomials, as they may not be intuitive. For example, for certain finite fields (which are rings), some formally different polynomials may be indistinguishable in terms of mappings.[5] Second, a polynomial may have more roots than its degree. Therefore, we will work in different rings R and provide conditions where our intuition is true in $R[x]$. It is clear that if you have two polynomials of degree n and m, their sum may be degree $k < n, m$. This is not always true for multiplication. There is a simple condition in which the degree is additive, however.

Lemma 8.1 (Bounds on Degrees From Operations)

Given that R is a ring and $f, g \in R[x]$,

$$\deg(f+g) \leq \max\{\deg f, \deg g\} \deg(fg) \leq \deg f \deg g \tag{418}$$

Proof. Trivial.

Unsurprisingly, the properties of the base ring R determines the properties on $R[x]$, and we will explore more of this later. Finally, let's view polynomials as functions $f : R \to R$, which will allow us to talk about their *roots*.

Definition 8.3 (Polynomial Root)

An element $r \in R$ is a **root** of polynomial $f \in R[x]$ if and only if

$$f(r) = 0 \tag{419}$$

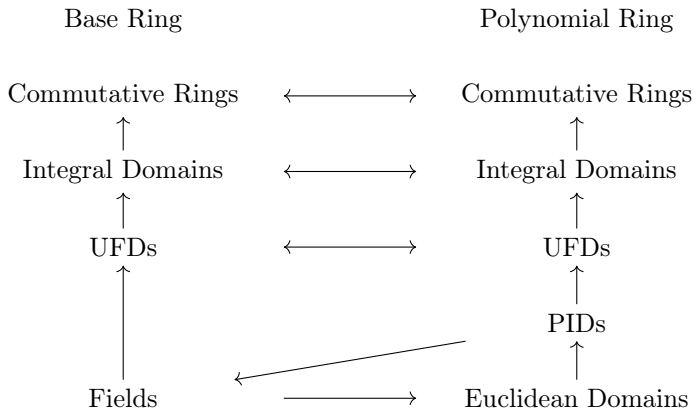


Figure 15: Basic hierarchy of rings with UFDs included.

8.1 Commutative Polynomial Rings

We begin our study by looking at commutative polynomial rings. The first order of business is to be able to tell when such $R[x]$ is commutative. There is a simple condition for this.

Theorem 8.2 (Conditions for $R[x]$ to be Commutative)

Let R be a ring and $S = \{x_1, \dots, x_m\}$ be a finite set of indeterminates. R is a commutative ring iff $R[x]$ is a commutative ring.

[5] x and x^2 are equivalent in $\mathbb{Z}_2[x]$.

Proof. We prove bidirectionally.
1. ($\rightarrow$) If R is commutative then given $a, b \in R$, we have $(ax^i)(bx^j) = a(x^ib)x^j = (ab)(x^ix^j) = (ba)(x^jx^i) = (bx^j)(ax^i)$, and so from distributive property $R[x]$ is commutative.
2. ($\leftarrow$) This is trivial since given $R[x]$ commutative, take $a, b \in R \subset R[x]$, and so $ab = ba$ in $R[x]$ implies commutativity in R.

Now we associate long division with Euclidean domains, and this is the case in the integers. However, for polynomials, we can state a slightly weaker form in that we can divide any polynomial by not just any other polynomial, but a *monic* polynomial.

Theorem 8.3 (Division Algorithm Exists for Monic Divisors)

Given any ring R, with $f(x), d(x) \in R[x]$ where the leading coefficient of $d(x)$ is a unit, a division algorithm exists. That is, we can find a $q(x), r(x) \in R$ where

$$f(x) = q(x)d(x) + r(x), \qquad 0 \leq \deg r(x) < \deg d(x) \tag{420}$$

Proof. We construct how to do so. This is because at each step, you only need to divide the leading coefficient of the divisor into the leading coefficient of the polynomial you have left.

Example 8.1 (Polynomial Long Division)

We apply the steps in the proof above to the example.

$$\begin{array}{rl} & x^2 + 6x + 11 \\ \hline x-2 \big) & x^3 + 4x^2 - x + 7 \\ & \underline{-x^3 + 2x^2} \\ & 6x^2 - x \\ & \underline{-6x^2 + 12x} \\ & 11x + 7 \\ & \underline{-11x + 22} \\ & 29 \end{array}$$

Now let's introduce polynomial roots and how they relate to the factors and evaluation of a function $f(x)$.

Corollary 8.4 (Root-Factor Theorem)

Given a commutative ring R and $f(x) \in R[x]$, $(x - c)$ is a factor of $f(x)$, i.e. can be factored into

$$f(x) = (x - c)q(x) \tag{421}$$

for some $q(x) \in R[x]$ of degree $\deg(f) - 1$ if and only if $f(c) = 0$.[a]

[a] Note that this is not true for an arbitrary ring. R must be commutative at least.

Proof. We prove for when R is a field F, but it turns out that the theorem also holds for commutative rings R.
1. ($\rightarrow$). Given that $(x - c)$ is a factor of $f(x)$, this means that by the Euclidean algorithm $f(x) = (x - c)q(x)$ for some $q(x)$, and so $f(c) = (c - c)q(c) = 0$.
2. ($\leftarrow$). Given that $f(c) = 0$. By the remainder theorem this means that when we divide $f(x)$ by $(x - c)$, the remainder is $f(c) = 0$, and so $f(x) = (x - c)q(x) + 0 = (x - c)q(x) \implies (x - c)$ is a

factor of $f(x)$.

Corollary 8.5 (Remainder Theorem)

Let $c \in F$ and $f(x) \in F[x]$. When we divide $f(x)$ by $g(x) = x - c$, the remainder is $f(c)$.

Proof. By the division algorithm,

$$f(x) = (x-c)q(x) + r(x) \implies f(c) = (c-c)q(c) + r(c) = r(c) \tag{422}$$

Definition 8.4 (Multiplicity)

A root c of polynomial $f(x) \in F[x]$ is called simple if $f(x)$ is not divisible by $(x-c)^2$ and multiple otherwise. The **multiplicity** of a root c is the maximum k such that $(x-c)^k$ divides $f(x)$.

Often, we refer to a polynomial having no repeated roots as being *square-free*.

Finally, let's talk about what quotient polynomial rings look like.

Example 8.2 (Counting Ring Homomorphisms)

We wish to count the number of ring homomorphisms

$$\phi : \frac{\mathbb{Z}}{\langle x^3 + y^2 - 1 \rangle} \to \mathbb{Z}_7 \tag{423}$$

Note that any such homomorphism ϕ induces by composition a unique canonical homomorphism $f : \mathbb{Z}[x,y] \to \mathbb{Z}_7$. f must map 1 to 1, so it leaves a total of 49 choices for what it can map x, y to. Now since the kernel must contain the ideal $\langle x^3 + y^2 - 1 \rangle$, this means that we would like

$$f(x^3 + y^2 - 1) = f(x)^3 + f(y)^2 - 1 = 0 \tag{424}$$

and from this we can count.

8.2 Polynomial Integral Domains

Now we talk about when $R[x]$ becomes an integral domain. When is this the case?

Theorem 8.6 (Conditions for $R[x]$ to be Integral Domain)

Let R be a ring and $S = \{x_1, \dots, x_m\}$ be a finite set of indeterminates. R is an integral domain iff $R[x]$ is an integral domain.

Proof. It suffices to prove the domain property since commutativity is already proved.

1. $(\to)$. Assume R is a domain. Now take any two nonzero polynomials $f(x), g(x) \in R[x]$. Then look at their leading term ax^n and bx^n. The leading coefficient of $(fg)(x)$ is $(ab)x^{n+m}$, and since R is a domain $ab \neq= 0 \implies (fg)(x) \neq 0$. So $R[x]$ is a domain.
2. $(\leftarrow)$. This is trivial since given $R[x]$ integral domain, take $a, b \in R \subset R[x]$ with $a, b \neq 0$, and so $ab \neq 0$ since $R[x]$ is a domain. Therefore R is a domain.

We can see how $R[x]$ fails to be a domain when R is not an integral domain.

Example 8.3 (Product of Two Linear Polynomials is 0)

Given $f, g \in \mathbb{Z}_6[x]$ with $f(x) = 2x + 4$ and $g(x) = 3x + 3$, we have

$$f(x) \cdot g(x) = (2x + 4)(3x + 3) = 6x^2 + 18x + 12 = 0 \tag{425}$$

In addition to the results we derived for general integral domains, what special results can we derive for that of polynomial rings? First, we have the familiar property of the degree of a product of two polynomials.

Lemma 8.7 (Degree of Product of Polynomials)

Given integral domain $R[x]$ and $f(x), g(x) \in R[x]$, we have

$$\deg fg(x) = \deg f(x) + \deg g(x) \tag{426}$$

Proof. Trivial. Consider the leading coefficient of $f(x), g(x)$.

Now the integral domain property gives us also a nice bound on the number of roots. If $R[x]$ was not an integral domain, then R is not an integral domain, and so there exists zero divisors $a, b \in R$ s.t. $ab = 0$. Now consider

$$f(x) = (x - a)(x - b) = x^2 - (a + b)x + ab = x^2 - (a + b)x \tag{427}$$

and we have found a degree two polynomial having at least three roots. We can also state the converse of this.

Theorem 8.8 (Bounds on Number of Roots)

Let $R[x]$ be an integral domain and $f(x) \in R[x]$. Then $f(x)$ has at most $\deg f(x)$ roots.

Proof. Let us consider the equation $f(x) = x^2 - m = 0$, where $m \in R$ is nonzero. Suppose $f(x)$ has more than 2 roots, i.e.

$$x^2 - m = (x - a)(x - b) = (x - c)(x - d), \qquad a, b, c, d \in R \tag{428}$$

Let $c \neq a, b$. Then $(c - a)(c - b) = 0(c - d) = 0$, which implies that R is not an integral domain.

Therefore, the number of roots of a polynomial—counted with multiplicity—does not exceed the degree of this polynomial. Furthermore, these numbers are equal if and only if the polynomial is a product of linear factors, which has its own terminology. That is, given a polynomial $f(x)$, we can view it as an element of multiple polynomial rings $R[x]$. The properties of the ring R will determine the properties of $f(x)$ as an element of $R[x]$.

Theorem 8.9 (Splitting Ring/Field)

A polynomial $f(x)$ is said to **split** in $R[x]$ if $f(x)$ can be factored into only linear factors.

8.3 Polynomial Unique Factorization Domains

What's great about UFDs R is that we can relate the decomposition of polynomials in $R[x]$ to the decomposition in $F[x]$, where F is the field of fractions of R.[6] This is precisely stated through Gauss's lemma.

[6] It is a field since R is an integral domain.

Lemma 8.10 (Gauss's Lemma)

Let R be a UFD and F be its field of fractions. Then reducibility in $F[x]$ implies reducibility in $R[x]$. That is, given $f(x) \in R[x]$, if there exists $g(x), h(x) \in F[x]$ s.t. $g(x)h(x) = f(x)$, then there exists $\bar{g}(x), \bar{h}(x) \in R[x]$ s.t. $\bar{g}(x)\bar{h}(x) = f(x)$.

Proof. We prove for $R = \mathbb{Z}$. We can find $k, l \in \mathbb{Z}$ s.t. $g_1(x) = kg(x)$ and $h_1(x) = lh(x)$ have integer coefficients, i.e. $g_1, h_1 \in \mathbb{Z}[x]$. Then, $klf(x) = g_1(x)h_1(x) \in \mathbb{Z}[x]$. Let p be a prime factor of kl. We have

$$0 \equiv \bar{k}\bar{l}\bar{f}(x) \equiv \bar{g}_1(x)\bar{h}_1(x) \text{ in } \mathbb{Z}_p[x] \tag{429}$$

Since $\mathbb{Z}_p$ is an integral domain, $\mathbb{Z}_p[x]$ is an integral domain, and so $\bar{g}_1$ or $\bar{h}_1$ must be 0. WLOG let it be $\bar{g}_1$. Then every coefficient of $g_1(x)$ is divisible by p, and we can write it in the form $g_2(x) = pg_1(x)$. Therefore,

$$p(x) \cdot \frac{kl}{p} = \underbrace{\frac{g_1(x)}{p}}_{g_2(x)} \cdot \underbrace{h_1(x)}_{h_2(x)} \iff f(x)\frac{kl}{p} = g_2(x)h_2(x) \tag{430}$$

Since there are only finitely many prime divisors, we do this for all prime factors of kl, and we have

$$f(x) = g_n(x)h_n(x), \qquad g_n, h_n \in \mathbb{Z}[x] \tag{431}$$

Therefore in the specific case of $\mathbb{Z}$, Gauss's lemma says that decompositions in $\mathbb{Q}[x]$ imply decompositions in $\mathbb{Z}[x]$! By looking at the contrapositive, to check irreducibility in $\mathbb{Q}[x]$, it suffices to check irreducibility in $\mathbb{Z}[x]$.

Theorem 8.11 (Conditions for $R[x]$ to be a UFD)

Let R be a ring and $S = \{x_1, \dots, x_m\}$ be a finite set of indeterminates. R is a UFD implies $R[S]$ is a UFD.

Proof. By induction, it suffices to show for when $S = \{x\}$. Let F be the field of fractions of R. Since F is a field, $F[x]$ is Euclidean domain and hence a PID. So every polynomial in $R[x]$ has a unique factorization in $F[x]$. By Gauss's lemma, this factorization is actually in $R[x]$.

Note that this is *not* true in arbitrary rings, i.e. non-UFDs.

Example 8.4 (Linear Polynomial with 3 Roots)

Consider $f(x) = x^2 - 1 \in \mathbb{Z}_8[x]$, a commutative ring. Then $1, 3, 5, 7$ are all roots of $f(x)$, which is greater than its degree. Furthermore, it has two different factorizations

$$x^2 - 1 = (x + 1)(x - 1) = (x + 3)(x - 3) \tag{432}$$

Another milestone theorem is that in UFDs, we can use the rational root theorem to reduce our search space of roots in the field of fractions.

Theorem 8.12 (Rational Root Theorem for UFDs)

Let R be a UFD and

$$f(x) = a_n x^n + \ldots + a_1 x + a_0 \in R[x] \tag{433}$$

If K is the fraction field of R and $f(k) = 0$ with $k = \frac{p}{q}$ for p, q coprime, then $p \mid a_0$ and $p \mid a_n$.

Proof. Given that r/s is a root, we have

$$a_n(r/s)^n + \ldots + a_0 = 0 \tag{434}$$

Multiplying by s^n, we get

$$a_n r^n + a_{n-1} r^{n-1} s + \ldots + a_1 s^{n-1} r + a_0 s^n = 0 \tag{435}$$

and putting this equation on mod r and mod s implies that $r | a_0 s^n$ and $s | a_n r^n$, respectively. But since we assumed that $\gcd(r, s) = 1$, $r | a_0$ and $s | a_n$.

Corollary 8.13 (Rational Root Theorem for Integers)

Let $a_n x^n + \ldots + a_0 \in \mathbb{Z}[x]$. If $r/s \in \mathbb{Q}$ with $\gcd(r, s) = 1$, then $r \mid a_0$ and $s \mid a_n$.

Example 8.5 (Reducibility of Integer Polynomials)

Let $f(x) = x^4 - x^3 + 2$. The rational roots are in the set $S = \{\pm 1, \pm 2\}$, but none of them work since $f(\pm 1), f(\pm 2) \neq 0$. By degree considerations and Gauss's lemma, if $f(x)$ is reducible, then

$$f(x) = (x^2 + ax + b)(x^2 + cx + d), \qquad a, b, c, d \in \mathbb{Z} \tag{436}$$

We know that $bd \in S$, with $a + c = -1$, $d + b + ac = 0$, and so on for each coefficients. We can brute force this finite set of possibilities.

Example 8.6

We claim that $x^4 - 22x^2 + 1 = (x - (\sqrt{6} + \sqrt{5}))(x - (\sqrt{6} - \sqrt{5}))(x - (-\sqrt{6} + \sqrt{5}))(x - (-\sqrt{6} - \sqrt{5}))$ is irreducible in $\mathbb{Q}[x]$. By the rational root theorem ± 1 are the possible rational roots, but plugging it into reveals that they aren't roots. Now if $x^4 - 22x^2 + 1$ factors it must factor as a product of monic quadratic polynomials in $\mathbb{Z}[x]$ by Gauss's Lemma. Therefore,

$$(x^2 + ax + b)(x^2 + cx + d) = x^4 - 22x^2 + 1 \tag{437}$$

Then $a + c = 0, bd = 1, ad + bc = 0, d + b + ac = 22$. We can derive this and find that there are no solutions, and so it is irreducible.

A great way to check irreducibility is to check in mod p.

Theorem 8.14

Let $f(x) = a_n x^n + \ldots + a_0 \in \mathbb{Z}[x]$. If $p \nmid a_n$ and $f \in \mathbb{Z}_p[x]$ is irreducible, then f is irreducible in $\mathbb{Q}[x]$.[a]

[a] May need to verify this again.

Proof. Suppose that $f(x) = g(x)h(x) \in \mathbb{Z}[x]$ with $\deg(g), \deg(h) > 0$. Then

$$f(x) \equiv g(x)h(x) \text{ in } \mathbb{Z}_p[x] \tag{438}$$

Since $f(x)$ is irreducible in $\mathbb{Z}_p[x]$, we must have that one of $g(x)$ or $h(x)$ has degree 0 in $\mathbb{Z}_p[x]$. WLOG

let it be $g(x)$, but this means that the leading coefficient of $g(x)$ must be divisible by $p \implies$ leading coefficient of $f(x)$ is divisible by $p \iff p \mid a_n$.

Example 8.7

x^4+x+1 is irreducible in $\mathbb{Z}_2[x]$. So we can extend this to $\mathbb{Z}[x]$ to see that *all* fourth degree polynomials of form $ax^4 + bx^3 + cx^2 + dx + e$, which a, d, e odd and b, c even is irreducible in $\mathbb{Q}[x]$.

This is a powerful theorem to quickly find a large class of polynomials that are irreducible. However, being reducible in $\mathbb{Z}_p[x]$ does not imply reducibility in $\mathbb{Q}$. In fact, there are polynomials $f(x) \in \mathbb{Z}[x]$ which are irreducible but reducible in $\mathbb{Z}_p$ for *every* prime p.

Theorem 8.15 (Eisenstein's Criterion)

Let $f(x) = a_n x^n + \ldots + a_0 \in \mathbb{Z}[x]$ and $p \in \mathbb{Z}$ a prime s.t. $p \nmid a_n$, $p \mid a_i$ for $i = 0, \ldots, a_{n-1}$, and $p^2 \nmid a_0$. Then $f(x)$ is irreducible in $\mathbb{Q}[x]$.

Proof. Suppose that $f(x) = g(x)h(x) \in \mathbb{Q}[x]$ with $\deg(g), \deg(h) > 0$. Then, by Gauss's lemma, $g, h \in \mathbb{Z}[x]$. Reducing the equations mod p,

$$f(x) = g(x)h(x) \text{ in } \mathbb{Z}_p[x] \tag{439}$$

But $f(x) = a_n x^n$. By unique factorization theorem in $\mathbb{Z}_p[x]$, $g, h \in \mathbb{Z}_p[x]$ must be products of units and prime factors of $a_n x^n$, which are $\{x\}$. Therefore, let

$$g(x) = b_m x^m, h(x) = \frac{a_n}{b_m} x^{n-m} \in \mathbb{Z}_p[x] \tag{440}$$

with $\deg(g) = m > 0$ and $\deg(h) = n - m > 0$ in $\mathbb{Z}[x]$. This implies that the constant coefficients of $g(x), h(x)$ are divisible by p, which implies that the constant coefficients of $f(x) = g(x)h(x)$ are divisible by p^2, a contradiction.

Example 8.8 (Easy Checks for Irreducibility with Eisenstein)

Listed.

1. $x^{13} + 2x^{10} + 4x + 6$ is irreducible in $\mathbb{Q}[x]$ by Eisenstein for $p = 2$.
2. $x^3 + 9x^2 + 12x + 3$ is irreducible in $\mathbb{Q}[x]$ by Eisenstein for $p = 3$.
3. Let $f(x) = x^4 + x^3 + x^2 + x + 1$. Then, we know that $f(x) = \frac{x^5-1}{x-1}$ and so

$$f(x+1) = \frac{(x+1)^5 - 1}{(x+1) - 1} \tag{441}$$

$$= \frac{1}{x}\left(x^5 + \binom{5}{1}x^4 + \binom{5}{2}x^3 + \binom{5}{3}x^2 + \binom{5}{4}x + \binom{5}{5} - 1\right) \tag{442}$$

$$= x^4 + 5x^3 + 10x^2 + 10x + 5 \tag{443}$$

So all nonleading coefficients are divisible by 5 exactly once, which by Eisenstein implies that $f(x+1)$ is irreducible which implies that $f(x)$ is irreducible.

Since UFDs have unique GCDs, we can state the following.

Theorem 8.16 (GCD of Two Polynomials in a UFD Exist)

Given nonzero polynomials $f(x), g(x) \in F[x]$, let

$$S = \{h(x) \in F[x] \mid h(x) = a(x)f(x) + b(x)g(x) \text{ for some } a(x), b(x) \in F[x]\} \tag{444}$$

Then there exists some polynomial $d(x) \in S$ of smallest degree, and every $h(x) \in S$ is divisible by $d(x)$.

Proof. The existence is trivial since by the well-ordering principle on the degrees of polynomials in S, such a minimal degree must exist. Now we prove the second claim by proving $d(x) \mid f(x)$. We apply the division algorithm to write

$$f(x) = q(x)d(x) + r(x) \tag{445}$$

If $r(x) = 0$, then by root factor theorem we are done. If $r(x) \neq 0$, we then write

$$\begin{aligned} r(x) &= f(x) - q(x)d(x) && (446)\\ &= f(x) - \big(s(x)f(x) + t(x)g(x)\big)q(x) && (447)\\ &= \big(1 - s(x)q(x)\big)f(x) - \big(t(x)q(x)\big)g(x) \in S && (448) \end{aligned}$$

Since $r(x) \in S$ due to its form, the fact that $\deg(r(x)) < \deg(d(x))$ contradicts the way that $d(x)$ was chosen. Therefore $r(x) = 0$. It turns out that $d(x)$ is unique up to a constant factor.

The algorithmic way for computing the GCD is done the same way by performing Euclidean algorithm on two polynomials: dividing on by the other, taking the remainder, and dividing the lesser degree by the remainder again, until the remainder is 0.

8.4 Polynomial Euclidean Domains

Definition 8.5 (Polynomial Euclidean Domains)

We call $F[x]$ a **Euclidean domain** if for all $f(x), g(x) \in F[x]$ and $g(x) \neq 0$, there exists polynomials $q(x), r(x)$ such that

$$f(x) = q(x)g(x) + r(x), \qquad 0 \leq \deg(r) < \deg(g) \tag{449}$$

where deg is the norm.

Example 8.9

We again start with the conditions for a polynomial ring to be a Euclidean domain. It is *not* the case that R should be a Euclidean domain. In fact, we need something stronger.

Theorem 8.17 (Conditions for $R[x]$ to be Euclidean Domain)

Let R be a ring and $S = \{x_1, \dots, x_m\}$ be a finite set of indeterminates. F is a field implies $R[x]$ is a Euclidean domain, with norm $N(f(x)) = \deg f(x)$. That is, given polynomials $f(x), g(x) \in F[x]$, there are unique polyomials $q(x), r(x) \in F[x]$ s.t.

$$f(x) = q(x)g(x) + r(x), \qquad \deg(r(x)) < \deg(g(x)) \tag{450}$$

Proof. We first prove existence. If $\deg(f(x)) < \deg(g(x))$, then we can trivially set $q(x) = 0, r(x) = f(x)$. Therefore we can assume that $\deg(f(x)) \geq \deg(g(x))$. We can prove this by strong induction on $k = \deg(f(x))$. Assume that $\deg(f(x)) = 1$. Then if $\deg(g(x)) > 1$ it is trivial as before, so we show for $\deg(g(x)) = 1$. So let

$$f(x) = a_1 x + a_0, \qquad g(x) = b_1 x + b_0 \tag{451}$$

and we can find the solutions

$$f(x) = \frac{a_1}{b_1} g(x) + \left(a_0 - \frac{a_1 b_0}{b_1} \right) \tag{452}$$

Now suppose that the results is known for whenver $\deg(f(x)) \leq k$ and we have a polynomial $F(x) = a_{k+1}x^{k+1} + \dots a_0$ of degree $k+1$. Then we must check that there exists a quotient and remainder for $0 \leq \deg(g(x)) = m \leq k+1$. Note that the coefficients of x^{k+1} in $F(x)$ and in the polynomial $\frac{a_{k+1}}{b_m} x^{k+1-m} g(x)$ are the same, so the polynomial

$$f(x) = F(x) - \frac{a_{k+1}}{b_m} x^{k+1-m} g(x) \tag{453}$$

has degree at most k. Thus by our induction hypothesis we can write $f(x) = q(x)g(x) + r(x)$, and so

$$\begin{aligned} F(x) &= f(x) + \frac{a_{k+1}}{b_m} x^{k+1-m} g(x) && (454) \\ &= q(x)g(x) + r(x) + \frac{a_{k+1}}{b_m} x^{k+1-m} g(x) && (455) \\ &= \left(q(x) + \frac{a_{k+1}}{b_m} x^{k+1-m} \right) g(x) + r(x) && (456) \end{aligned}$$

which is indeed a decomposition. Now to prove uniqueness, suppose we had two different decompositions

$$f(x) = q(x)g(x) + r(x) = q'(x)g(x) + r'(x) \implies \big(q(x) - q'(x)\big)g(x) = r(x) - r'(x) \tag{457}$$

IF $q(x) \neq q'(x)$, then the degree of the LHS is at least $\deg(g(x))$, while the degree of the RHS must be strictly less, a contradiction.

Note that alternatively, we have shown from 8.3 that the leading coefficient of the divisor $g(x)$ must be a unit in R. For a field, every element is a unit, and we are done.

Theorem 8.18 (PID Polynomial Rings Implies Underlying is a Field)

$R[x]$ is a PID implies R is a field.

Proof. Assume that $R[x]$ is a PID. Since R is a subring of $R[x]$, then R must be an integral domain. The ideal $\langle x \rangle$ is a nonzero prime ideal in $R[x]$ because $R[x]/\langle x \rangle$ is isomorphic to the integral domain R. Since every nonzero prime ideal in a PID is a maximal ideal, $\langle x \rangle$ is a maximal ideal, hence the quotient R is a field.

From this theorem, we can see that if a polynomial ring is a PID, then we know that it is automatically a Euclidean domain, so it suffices to only consider the case of Euclidean domains, which is why we skip PID polynomial rings.

Now we talk about the behavior of polynomials as functions. Sometimes, we may have two different polynomials but they may define the same function from R to R!

Example 8.10 (Polynomials as Same Function)

Given $f, g \in \mathbb{Z}_2[x]$,

$$f(x) = x \sim g(x) = x^2 \tag{458}$$

With an additional condition on a field, we can guarantee that each polynomial determines a different function.

Theorem 8.19 (Uniqueness of Polynomials over Field)

If the field $\mathbb{F}$ is infinite, then different polynomials in $\mathbb{F}[x]$ determine different functions.

Theorem 8.20 (Interpolation)

For any collection of given field values $y_1, y_2, ..., y_n \in F$ at given distinct points $x_1, x_2, ..., x_n \in F$, there exists a unique polynomial $f \in F[x]$ with $\deg f < n$ such that

$$f(x_i) = y_i, \quad i = 1, 2, ..., n \tag{459}$$

This is commonly known as the **interpolation problem**, and when $n = 2$, this is called **linear interpolation**.

8.5 Algebraically Closed Fields

Now that we have seen some examples of fields, what properties would we like it to have? Going back to polynomials, recall that if F is a field, then $F[x]$ as a Euclidean domain gave us a lot of nice properties, such as admitting a unique factorization of irreducible polynomials. However, we have only proved that the number of roots is *at most* the degree n, but not that it actually reaches n. In fact, in a more extreme case, a polynomial may not even factor *at all* in $F[x]$, since it could be irreducible. So while we have defined an upper bound for the number of roots for a polynomial, we have not determined whether a polynomial has any roots at all, i.e. a lower bound.

We don't have much *control* over what these irreducible polynomials can look like. We may have to check—either through theorems or manually—that a polynomial or arbitrary degree is irreducible. If we would like to assert that all irreducible polynomials must be of smallest degree—that is, linear—then such a field is called *algebraically closed*. This algebraic closed property asserts also that the lower bound on the number of (non-unique) factors is n.

Definition 8.6 (Algebraically Closed Field)

A field F is **algebraically closed** if every polynomial of positive degree (i.e. non-constant) in $F[x]$ has at least one root in F.

This is equivalent to saying that every polynomial can be expressed as a product of first degree polynomials. To extend our analysis more, we can talk about the multiplicity of these factors, which just tells us more about how many unique and non-unique factors a polynomial has.

Example 8.11 (Reals are not Algebraically Closed)

$\mathbb{R}$ is not algebraically closed since we can identify the polynomial $f(x) = x^2 + 1 \in \mathbb{R}[x]$ which does not have any roots in $\mathbb{R}$. Consequently, any subfield of $\mathbb{R}$ (which contains 1) such as $\mathbb{Q}, \mathbb{Q}(\sqrt{2}), \dots$ are not algebraically closed.

It turns out that the complex numbers are algebraically closed, which is presented with the following grand name. Ironically, this theorem cannot be proven with algebra alone. We need complex analysis.[7]

Theorem 8.21 (Fundamental Theorem of Algebra)

Suppose $f \in \mathbb{C}[x]$ is a polynomial of degree $n \geq 1$. Then $f(x)$ has a root in $\mathbb{C}$. It immediately follows from induction that it can be factored as a product of linear polynomials in $\mathbb{C}[x]$.

Proof. WLOG we can assume that f is monic: $f(z) = z^n + a_{n-1}z^{n-1} + \ldots + a_1 z + a_0$. Since $\mathbb{C}$ is a field, we can set

$$f(z) = z^n \left(1 + \frac{a_{n-1}}{z} + \frac{a_{n-2}}{z^2} + \ldots + \frac{a_0}{z_n}\right) \tag{460}$$

Since

$$\lim_{|z| \to \infty} \left(1 + \frac{a_{n-1}}{z} + \frac{a_{n-2}}{z^2} + \ldots + \frac{a_0}{z_n}\right) = 0 \tag{461}$$

there exists a $R > 0$ s.t.

$$|z| > R \implies \left|1 + \frac{a_{n-1}}{z} + \frac{a_{n-2}}{z^2} + \ldots + \frac{a_0}{z_n}\right| < \frac{1}{2} \tag{462}$$

and hence

$$|z| > R \implies |f(z)| > |z|^n \cdot \left(1 - \frac{1}{2}\right) > \frac{R^n}{2} \tag{463}$$

So z cannot be a root if $|z| > R$. On the other hand, $f(z)$ is continuous (under the Euclidean topology) and so on the compact set $\{z \in \mathbb{C} \mid |z| \leq R\}$, $|f(z)|$ achieves a minimum value say at the point z_0. We claim that $\min_z f(z) = 0$.

For convenience, we let $z_0 = 0$ (we can do a change of basis on the polynomial) and assume that the minimum is some positive number, i.e. $f(0) = a_0 \neq 0$. Let j be the smallest positive integer such that $a_j = 0$. Let

$$g(z) = \frac{a_{j+1}}{a_j} z + \ldots + \frac{a_n}{a_j} z^{n-j} \implies f(z) = a_0 + a_j z^j \big(1 + g(z)\big) \tag{464}$$

We set $\gamma = \sqrt[j]{-a_0/a_j}$ and consider the values of

$$\begin{aligned} f(t\gamma) &= a_0 + a_j (t\gamma)^j \big(1 + g(t\gamma)\big) && (465)\\ &= a_0 - a_0 t^j \big(1 + g(t\gamma)\big) && (466)\\ &= a_0 \big\{1 - t^j \big(1 + g(t\gamma)\big)\big\} && (467) \end{aligned}$$

for $t > 0$. For t sufficiently small, we have

$$|g(t\gamma)| = \left|\frac{a_{j+1}}{a_j}(t\gamma) + \ldots + \frac{a_n}{a_j}(t\gamma)^{n-j}\right| < \frac{1}{2} \tag{468}$$

and for such t, this implies

$$|f(t\gamma)| = |a_0||1 - t^j(1 + g(t\gamma))| \leq |a_0||1 - t^j/2| < |a_0| \tag{469}$$

and so z_0 cannot have been the minimum of $|f(z)|$. Therefore, the minimum value must be 0.

Great, so through this theorem, we can work in any subfield of $\mathbb{C}$ and guarantee that will have all of its roots in $\mathbb{C}$.

[7] Gauss proved this for the first time in 1799.

Corollary 8.22 ($\mathbb{C}$ is algebraically closed)

$\mathbb{C}$ is algebraically closed, i.e. $\mathbb{C}$ is a splitting field of $\mathbb{C}[x]$.

Put more succinctly, the impossibility of defining division on the ring of integers motivates its extension into the field of rational numbers. Similarly, the inability to take square roots of negative real numbers forces us to extend the field of real numbers to the bigger field of complex numbers.

Theorem 8.23 (Complex Roots Come in Conjugate Pairs)

If α is a complex root of polynomial $f \in \mathbb{R}[x]$, then $\bar{\alpha}$ is also a root of the polynomial. Moreover, $\bar{\alpha}$ has the same multiplicity as α.

Proof. Note that conjugation is an automorphism (in fact an $\mathbb{R}$-automorphism over $\mathbb{C}$). Therefore, if $\alpha \in \mathbb{C}$ is a root, then

$$\begin{aligned} f(\bar{\alpha}) &= a_n f(\bar{\alpha})^n + \dots a_1 f(\bar{\alpha}) + a_0 && (470)\\ &= a_n \overline{f(\alpha)}^n + \dots a_1 \overline{f(\alpha)} + a_0 && (471)\\ &= \bar{0} = 0 && (472) \end{aligned}$$

This theorem has a lot of consequences on the reducibility of polynomials in the reals.

Corollary 8.24 (Odd Degree Polynomials have At Least 1 Real Root)

Every polynomial $f \in \mathbb{R}[x]$ of odd degree has at least one real root.

Proof. Assume that $n = \deg f$ was odd and there were m roots, with m even. Then there are $n - m$ roots, which is an odd number of complex roots, which contradicts the previous theorem.

Proof. Just for fun, a proof using analysis is as such. Without loss of generality we can assume that the leading coefficient of f is positive. Then

$$\lim_{x \to +\infty} f(x) = +\infty, \ \lim_{x \to -\infty} f(x) = -\infty \tag{473}$$

By the intermediate value theorem, there must be some point where f equals 0.

Corollary 8.25 (Real Polynomials Factors Into Linear and Quadratic Terms)

Every nonzero polynomial in $\mathbb{R}[x]$ factors into a product of linear terms and quadratic terms with negative discriminants.

Example 8.12

$$\begin{aligned} x^5 - 1 &= (x-1)\left(x - \left(\cos\frac{2\pi}{5} + i\sin\frac{2\pi}{5}\right)\right)\left(x - \left(\cos\frac{2\pi}{5} - i\sin\frac{2\pi}{5}\right)\right) \\ &\quad \times \left(x - \left(\cos\frac{4\pi}{5} + i\sin\frac{4\pi}{5}\right)\right)\left(x - \left(\cos\frac{4\pi}{5} - i\sin\frac{4\pi}{5}\right)\right) \\ &= (x-1)\left(x^2 - \frac{\sqrt{5}-1}{2}x + 1\right)\left(x^2 + \frac{\sqrt{5}+1}{2}x + 1\right) \end{aligned}$$

Lemma 8.26 (Symmetry on Sign of Roots)

The number of positive roots of $f(x) \in \mathbb{R}[x]$ is the same as the number of negative roots of $f(-x) \in \mathbb{R}[x]$.

Theorem 8.27 (Descartes' Rule of Signs)

Let $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in \mathbb{R}[x]$. Let C_+ be the number of times the coefficients of $f(x)$ change signs (here we ignore the zero coefficients); let Z_+ be the number of positive roots of $f(x)$, counting multiplicities. Then $Z_+ \leq C_+$ and $Z_+ \equiv C_+ \pmod 2$. Moreover, if we set $g(x) = f(-x)$, let C_- be the number of times the coefficients of $g(x)$ change signs, and Z_- the number of negative roots of $f(x)$. Then $Z_- \leq C_-$ and $Z_- \equiv C_- \pmod 2$.

Example 8.13 (Easy Way to Find Number of Positive Roots)

Given $f(x) = x^5 + x^4 - x^2 - 1$,

1. We have $C_+ = 1$. By Descartes' rule of signs, it must be the case that $Z_+ \leq 1$ and $Z_+ \equiv 1 \pmod 2 \implies Z_+ = 1$.
2. Since $f(-x) = -x^5 + x^4 - x^2 - 1$, we have $C_- = 2$, so $Z_- = 0$ or 2. This is the best that we can do, though it turns out that it actually has 0 negative roots.[a]

[a] On the other hand, $x^5 + 3x^3 - x^2 - 1$ has 2 negative roots.

Note that if a polynomial has a multiple root but its coefficients are known only approximately (but with any degree of precision), then it is impossible to prove that the multiple roots exists because under any perturbation of the coefficients, however small, it may separate into simple roots or simply cease to exist. This fact leads to the "instability" of the Jordan Normal form because under any perturbation of the elements of a matrix A, the change may drastically affect the characteristic polynomial, hence affecting the geometric multiplicities of its eigenvectors.

8.6 The Field of Rational Functions

Given a field F, we have constructed the Euclidean domain $F[x]$. However, this is one step away from being a field. We mimick the construction of the rational numbers $\mathbb{Q}$ as a quotient space over $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ by taking $F[x] \times (F[x] \setminus \{0\})$ and putting a quotient on it.

Definition 8.7 (Rational Functions)

The **rational functions** are defined to be the field of quotients (really just 2-tuples) of the form

$$F(x) \coloneqq \left\{ \frac{f(x)}{g(x)} \;\middle|\; f(x), g(x) \in F[x], g(x) \neq 0 \right\} \tag{474}$$

where addition and multiplication is defined in the usual sense.

Theorem 8.28 (Partial Fractions Decomposition)

Let $f(x), g(x) \in F[x]$ where $\deg(f(x)) < \deg(g(x))$. If $g(x) = u(x)v(x)$ where u, v are relatively prime, then there are polynomials $a(x), b(x)$ with $\deg(a) < \deg(u), \deg(b) < \deg(v)$ s.t.

$$\frac{f(x)}{g(x)} = \frac{a(x)}{u(x)} + \frac{b(x)}{v(x)} \tag{475}$$

By induction, we can prove this for any finite set of irreducible polynomials.

Proof. We describe an algorithm to get this decomposition. There are polynomials $s(x), t(x)$ s.t. $1 = s(x)u(x) + t(x)v(x)$. Therefore,

$$\frac{f(x)}{u(x)v(x)} = \frac{f(x)t(x)}{u(x)} + \frac{f(x)s(x)}{v(x)} \tag{476}$$

and we can use the Euclidean algorithm to write

$$\frac{f(x)t(x)}{u(x)} = q(x) + \frac{a(x)}{u(x)}, \qquad \deg(a) < \deg(u) \tag{477}$$

$$\frac{f(x)s(x)}{v(x)} = q(x) + \frac{a(x)}{u(x)}, \qquad \deg(b) < \deg(v) \tag{478}$$

which implies

$$\frac{f(x)}{u(x)v(x)} = \frac{a(x)}{u(x)} + \frac{b(x)}{v(x)} \tag{479}$$

Example 8.14

Consider the rational function $\frac{x+3}{x^3(x-1)^2}$. Applying the Euclidean algorithm, we find that

$$1 = (3x^2 + 2x + 1)(x-1)^2 - (3x-4)x^3 \tag{480}$$

and so

$$\frac{x+3}{x^3(x-1)^2} = \frac{(x+3)(3x^2+2x+1)}{x^3} - \frac{(x+3)(3x-4)}{(x-1)^2} \tag{481}$$

$$= \frac{11x^2+7x+3}{x^3} + \frac{-11x+15}{(x-1)^2} \tag{482}$$

8.7 Exercises

9 Modules

Now we give a generalization of vector spaces in which the underlying field is replaced by a general ring.

Definition 9.1 (Module)

Given a ring R, a **left R-module** M is an abelian group $(M, +)$ and an operation $\cdot : R \times M \to M$—called *scalar multiplication*—such that for all $r, s \in R$ and $x, y \in M$, we have

1. $r(x + y) = rx + ry$.
2. $(r + s)x = rx + sx$.
3. $(rs) \cdot x = r(sx)$
4. $1 \cdot x = x$.

Note that the "left" refers to the ring elements appearing on the left $R \times M$, and the analogous definition for right modules can be made.

Before going into any examples, we introduce submodules, which are subsets of modules M that are themselves modules under the restricted operations.

Definition 9.2 (Submodule)

Given a R-module M, a **submodule** N of M is a subgroup which is closed under the action of ring elements, i.e. $rn \in N$ for all $r \in R, n \in N$.

In particular, if $R = F$ is a field, then modules and submodules are the same as vector spaces and subspaces—though we haven't formally defined them yet.

Example 9.1 (Modules)

Let R be a ring.

1. R is a submodule, where scalar multiplication $\cdot : R \times R$ is the same as the ring multiplication in R.
2. Given $n \in \mathbb{N}$, the set
$$R^n \coloneqq \{(a_1, \dots, a_n) \mid \forall i, a_i \in R\} \tag{483}$$
is an R-module with addition and scalar multiplication defined component-wise. This is called the **free-module of rank n over R**.

Theorem 9.1 (Submodule Criterion)

Let R be a ring and M an R-module. A subset $N \subset M$ is a submodule of M if and only if

1. $N \neq \emptyset$, and
2. $x + ry \in N$ for all $r \in R$ and $x, y \in N$.

9.1 Modules over a PID

9.2 Rational Canonical Form

9.3 Jordan Canonical Form

9.4 Exercises

10 Vector Spaces

Definition 10.1 (Vector Space)

A **vector space over a field** F consists of an abelian group $(V, +)$ and an operation called **scalar multiplication**

$$\cdot : F \times V \to V \tag{484}$$

such that for all $x, y \in V$ and $\lambda, \mu \in F$, we have
1. $\lambda \cdot (x + y) = \lambda \cdot x + \lambda \cdot y$
2. $(\lambda + \mu) \cdot x = \lambda \cdot x + \mu \cdot x$
3. $(\lambda\mu) \cdot x = \lambda \cdot (\mu \cdot x)$, which equals $(\mu\lambda) \cdot x = \mu \cdot (\lambda \cdot x)$ since F is commutative
4. $1 \cdot x = x$, where 1 is the unity of F

Definition 10.2

A **left R-module** M consists of an abelian group $(M, +)$ and an operation called **scalar multiplication**

$$\cdot : R \times M \longrightarrow M \tag{485}$$

such that for all $\lambda, \mu \in R$ and $x, y \in M$, we have
1. $\lambda \cdot (x + y) = \lambda \cdot x + \lambda \cdot y$
2. $(\lambda + \mu) \cdot x = \lambda \cdot x + \mu \cdot x$
3. $(\lambda\mu) \cdot x = \lambda \cdot (\mu \cdot x)$, not necessarily equaling $(\mu\lambda) \cdot x = \mu \cdot (\lambda \cdot x)$
4. $1 \cdot x = x$, where 1 is the unity of R

Note that a left R-module is a vector space if and only if R is a field.

Definition 10.3

A **right R-module** M is defined analogously to a left R-module, except that the scalar multiplication operation is defined

$$\cdot : M \times R \longrightarrow M \tag{486}$$

Definition 10.4

Let A be a vector space over a field F equipped with an additional binary operation

$$\times : A \times A \longrightarrow A \tag{487}$$

A is an **algebra over** F if the following identities hold for all $x, y, z \in A$ and all $\lambda, \mu \in F$.
1. Right distributivity. $(x + y) \times z = x \times z + y \times z$
2. Left distributivity. $z \times (x + y) = z \times x + z \times y$
3. Compatibility with scalars. $(\lambda \cdot x) \times (\mu \cdot y) = (\lambda\mu) \cdot (x \times y)$

Note that vector multiplication of an algebra does not need to be commutative.

Example 10.1

The set of all $n \times n$ matrices with matrix multiplication is a noncommutative, associative algebra. Similarly, the set of all linear endomorphisms of a vector space V with composition is a noncommutative, associative algebra.

Example 10.2

$\mathbb{R}^3$ equipped with the cross product is an algebra, where the cross product is **anticommutative**, that is $x \times y = -y \times x$. $\times$ is also nonassociative, but rather satisfies an alternative identity called the **Jacobi Identity**.

Example 10.3

The set of all polynomials defined on an interval $[a, b]$ is an infinite-dimensional subalgebra of the set of all functions $f : \mathbb{R} \longrightarrow \mathbb{R}$ defined on $[a, b]$.

Definition 10.5

Similar to division rings, a **division algebra** is an algebra where the operation of "division" defined as such: Given any $a \in A$, nonzero $b \in A$, there exists solutions to the equation

$$A = bx \tag{488}$$

that are unique. If we wish, we can distinguish left and right division to be the solutions of $A = bx$ and $A = xb$.

Definition 10.6

Here are examples of division algebras.

1. $\mathbb{R}$ is a 1-dimensional algebra over itself.
2. $\mathbb{C}$ is a 2-dimensional algebra over $\mathbb{R}$.
3. There exists no 3-dimensional algebra.
4. Quaternions forms a 4-dimensional algebra over $\mathbb{R}$.

Theorem 10.1 (Eigenvector Conditions for Algebraic Closedness)

A field F is algebraically closed if and only if for each natural number n, every endomorphism of F^n (that is, ever linear map from F^n to itself) has at least one eigenvector.

Proof. An endomorphism of F^n has an eigenvector if and only if its characteristic polynomial has some root. ($\rightarrow$) So, when F is algebraically closed, every characteristic polynomial, which is an element of $F[x]$, must have a root. ($\leftarrow$) Assume that every characteristic polynomial has some root, and let $p \in F[x]$. Dividing the polynomial by a scalar doesn't change its roots, so we can assume p to have leading coefficient 1. If $p(x) = a_0 + a_1 x + ... + x^n$, then we can identify matrix

$$A = \begin{pmatrix} 0 & 0 & ... & 0 & -a_0 \\ 1 & 0 & ... & 0 & -a_1 \\ 0 & 1 & ... & 0 & -a_2 \\ ... & ... & ... & ... & ... \\ 0 & 0 & ... & 1 & -a_{n-1} \end{pmatrix} \tag{489}$$

such that the characteristic polynomial of A is p.

10.1 Modules

Vector space but over a ring.

10.2 Algebras

Vector space with bilinear product.

10.3 The Algebra of Quaternions

Definition 10.7

The **quaternions** form an algebra of 4-dimensional vectors over $\mathbb{R}$, with elements of the form

$$(a, b, c, d) \equiv a + bi + cj + dk \tag{490}$$

where a is called the **scalar portion** and $bi + cj + dk$ is called the **vector/imaginary portion**. The algebra of quaternions is denoted $\mathbb{H}$, which stands for "Hamilton." $\mathbb{H}$ is a 4-dimensional associative normed division algebra over $\mathbb{R}$.

From looking at the multiplication table, we can see that multiplication in $\mathbb{H}$ is not commutative.

i	i	-1	k	-j
j	j	-k	-1	i
k	k	j	-i	-1

Note the identity

$$i^2 = j^2 = k^2 = -1 \tag{491}$$

The algebra of quaternions are in fact the first noncommutative algebra to be discovered!

Theorem 10.2

$\mathbb{H}$ and $\mathbb{C}$ are the only finite-dimensional divisions rings containing $\mathbb{R}$ as a proper subring.

Definition 10.8

The **quaternion group**, denoted Q_8 is a nonabelian group of order 8, isomorphic to a certain 8-element subset in $\mathbb{H}$ under multiplication. It's group presentation is

$$Q_8 = \langle \bar{e}, i, j, k \mid \bar{e}^2 = e, i^2 = j^2 = k^2 = ijk = \bar{e} \rangle \tag{492}$$

Going back to the algebra, we can set $\{1, i, j, k\}$ as a basis and define addition and scalar multiplication component-wise, and multiplication (called the **Hamilton product**) with properties

1. The real quaternion 1 is the identity element.
2. All real quaternions commute with quaternions: $aq = qa$ for all $a \in \mathbb{R}, q \in \mathbb{H}$.
3. Every quaternion has an inverse with respect to the Hamilton product.

$$(a + bi + cj + dk)^{-1} = \frac{1}{a^2 + b^2 + c^2 + d^2}\big(a - bi - cj - dk\big) \tag{493}$$

Note that property 3 allows $\mathbb{H}$ to be a division algebra.

Theorem 10.3 (Scalar and Vector Components)

Let the quaternion be divided up into a scalar and vector part with the bjective mapping $a + bi + cj + dk \mapsto \big(a, (b, c, d)\big)$.

$$q = (r, v), r \in \mathbb{R}, v \in \mathbb{R}^3 \tag{494}$$

Then, the formulas for addition and multiplication are

$$\begin{aligned} q_1 + q_2 &= (r_1, v_1) + (r_2, v_2) = (r_1 + r_2, v_1 + v_2) \\ q_1 \cdot q_2 &= (r_1, v_1) \cdot (r_2, v_2) = (r_1 r_2 - v_1 \cdot v_2, r_1 v_2 + r_2 v_1 + v_1 \times v_2) \end{aligned}$$

where the $\cdot$ and $\times$ on the right hand side represnts the dot product and cross product, respectively.

Definition 10.9

The conjugate of a quaternion $q = a + bi + cj + dk$ is defined

$$\bar{q}, q^* \equiv a - bi - cj - dk \tag{495}$$

It has properties
1. $q^{**} = q$
2. $(qp)^* = p^* q^*$

q^* can also be expressed in terms of addition and multiplication.

$$q^* = -\frac{1}{2}\big(q + iqi + jqj + kqk\big) \tag{496}$$

Definition 10.10

The **norm** of q is defined

$$||q|| \equiv \sqrt{q^* q} = \sqrt{qq^*} = \sqrt{a^2 + b^2 + c^2 + d^2} \tag{497}$$

with properties
1. Scaling factor. $||\alpha q|| = |\alpha| ||q||$
2. Multiplicative. $||pq|| = ||p|| ||q||$

The norm allows us to define a metric

$$d(p, q) \equiv ||p - q|| \tag{498}$$

This makes $\mathbb{H}$ a metric space, with addition and multiplication continuous on the metric topology.

Definition 10.11

The **unit quaternion** is defined to be

$$U_q = \frac{q}{||q||} \tag{499}$$

Corollary 10.4

Every quaternion has a polar decomposition

$$q = U_q \cdot ||q|| \tag{500}$$

With this, we can redefine the inverse as

$$q^{-1} = \frac{q^*}{||q||^2} \tag{501}$$

10.3.1 Matrix Representations of Quaternions

We can represent q with 2×2 matrices over $\mathbb{C}$ or 4×4 matrices over $\mathbb{R}$.

Theorem 10.5

The following representation is an injective homomorphism $\rho : \mathbb{H} \longrightarrow \text{GL}(2, \mathbb{C})$.

$$\rho : a + bi + cj + dk \mapsto \begin{pmatrix} a+bi & c+di \\ -c+di & a-bi \end{pmatrix} \tag{502}$$

It has properties

1. Constraining any two of b, c, d to 0 produces a representation of the complex numbers. When $c = d = 0$, this is called the **diagonal representation**.

$$\begin{pmatrix} a+bi & 0 \\ 0 & a-bi \end{pmatrix}, \begin{pmatrix} a & c \\ -c & a \end{pmatrix}, \begin{pmatrix} a & di \\ di & a \end{pmatrix}$$

2. The norm of a quaternion is the square root of the determinant of its corresponding matrix representation.

$$||q|| = \sqrt{\det \begin{pmatrix} a+bi & c+di \\ -c+di & a-bi \end{pmatrix}} = \sqrt{(a^2+b^2)+(c^2+d^2)} \tag{503}$$

3. The conjugate of a quaternion corresponds to the conjugate (Hermitian) transpose of its matrix representation.

$$\rho(q^*) = \rho(q)^H \iff a - bi - cj - dk \mapsto \begin{pmatrix} a-bi & -c-di \\ c-di & a+bi \end{pmatrix} \tag{504}$$

4. The restriction of this representation to only unit quaternions leads to an isomorphism between the subgroup of unit quaternions and their corresponding image in SU(2). Topologically, the unit quaternions is the 3-sphere, so the underlying space SU(2) is also a 3-sphere. More specifically,

$$\frac{\text{SU}(2)}{2} \simeq \text{SO}(3) \tag{505}$$

Theorem 10.6

The following representation of $\mathbb{H}$ is an injective homomorphism $\rho : \mathbb{H} \longrightarrow \text{GL}(4, \mathbb{R})$.

$$\rho : a + bi + cj + dk \mapsto \begin{pmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{pmatrix} \tag{506}$$

or also as

$$a \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + b \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} + c \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} + d \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \tag{507}$$

It has properties

1. $\rho(q^*) = \rho(q)^T$
2. The fourth power of the norm is the determinant of the matrix

$$||q||^4 = \det\big(\rho(q)\big) \tag{508}$$

3. Similarly, with the 2×2 representation, complex number representations can be produced by restricting 2 of b, c, d to 0.

Note that this representation in $\text{GL}(4, \mathbb{R})$ is not unique. There are in fact 48 distinct representation of this form where one of the component matrices represents the scalar part and the other 3 are skew symmetric.

10.3.2 Square Roots of -1

In $\mathbb{C}$, there are two numbers, i and $-i$, whose square is -1. However, in $\mathbb{H}$, infinitely many square roots of -1 exist, forming the unit sphere in $\mathbb{R}^3$. To see this, let $q = a + bi + cj + dk$ be a quaternion, and assume that its square is -1. Then this implies that

$$a^2 - b^2 - c^2 - d^2 = -1, 2ab = 2ac = 2ad = 0 \tag{509}$$

To satisfy the second equation, either $a = 0$ or $b = c = d = 0$. The latter is impossible since then q would be real. Therefore,

$$b^2 + c^2 + d^2 = 1 \tag{510}$$

which forms the unit sphere in $\mathbb{R}^3$.

10.4 Tensor Algebras

Remember that an algebra is (loosely) a vector space V with a multiplication operation

$$\times : V \times V \longrightarrow V \tag{511}$$

Definition 10.12

The **tensor algebra** of vector space V over field $\mathbb{F}$ is

$$\begin{aligned} T(V) \equiv \bigoplus_{n=0}^{\infty} V^{\otimes n} &= V^{\otimes 0} \oplus V^{\otimes 1} \oplus V^{\otimes 2} \oplus V^{\otimes 3} \oplus ... \\ &= \mathbb{F} \oplus V \oplus V^{\otimes 2} \oplus V^{\otimes 3} \oplus V^{\otimes 4} \oplus ... \end{aligned}$$

with elements being infinite-tuples

$$(a, B^{\mu}, C^{\nu\gamma}, D^{\alpha\beta\epsilon}, ...) \tag{512}$$

The addition operation is defined component-wise, and the multiplication operation is the tensor product

$$\otimes : T(V) \times T(V) \longrightarrow T(V) \tag{513}$$

and the identity element is

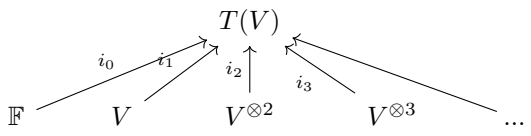
$$I = (1, 0, 0, ...) \tag{514}$$

Linearity can be easily shown.

The tensor algebra is often used to "add" differently ranked tensors together. But in order to do this rigorously, we must define the canonical injections

$$i_j : V^{\otimes j} \longrightarrow T(V),\ i_j(T^{\kappa_1,...,\kappa j}) = (0, ..., 0, T^{\kappa_1,...,\kappa j}, 0, ..., 0) \tag{515}$$

shown in the diagram



Therefore, with these i_j's, we can implicitly define the addition of arbitrary tensors $A \in V^{\otimes n}$ and $B \in V^{\otimes m}$ as

$$A + B \equiv i_n(A) + i_m(B) \in T(V) \tag{516}$$

along with multiplication of tensors as

$$A \otimes B \equiv i_n(A) \otimes i_m(B) \equiv i_{n+m}(A \otimes B) \tag{517}$$

We can also redefine the tensor product operation between two spaces to be an operation within $T(V)$ itself.

$$i_i(V^{\otimes i}) \otimes i_j(V^{\otimes j}) = i_{i+j}(V^{\otimes (i+j)}) \tag{518}$$

We can now proceed to define Exterior and Symmetric algebras as quotient algebras.

Definition 10.13

The **exterior algebra** $\Lambda(V)$ of a vector space V over field $\mathbb{F}$ is the quotient algebra of the tensor algebra $T(V)$

$$\Lambda(V) \equiv \frac{T(V)}{I} \tag{519}$$

where I is the two-sided ideal generated by all elements of the form $x \otimes x$ for $x \in V$ (i.e. all tensors that can be expressed as the tensor product of a vector in V by itself).
The **exterior product** $\wedge$ of two elements of $\Lambda(V)$ is the product induced by the tensor product $\otimes$ of $T(V)$. That is, if

$$\pi : T(V) \longrightarrow \Lambda(V) \tag{520}$$

is the canonical projection/surjection and $a, b \in \Lambda(V)$,then there are $\alpha, \beta \in T(V)$ such that $a = \pi(\alpha), b = \pi(\beta)$, and

$$a \wedge b = \pi(\alpha \otimes \beta) \tag{521}$$

We can define this quotient space with the equivalence class

$$x \otimes y = -y \otimes x \pmod{I} \tag{522}$$

Definition 10.14

The **symmetric algebra** $\text{Sym}(V)$ of a vector space V over a field $\mathbb{F}$ is the quotient algebra of the tensor algebra $T(V)$

$$\Lambda(V) \equiv \frac{T(V)}{J} \tag{523}$$

where J is the two-sided ideal generated by all elements in the form

$$v \otimes w - w \otimes v \tag{524}$$

(i.e. commutators of all possible pairs of vectors).

10.5 Exercises

11 Galois Theory

In general, given a polynomial $f(x)$, we would like to find effective ways to factor it or find its roots. This depends on which polynomial ring $R[x]$ that we view $f(x)$ as an element of, and finding such a unique factorization requires us to work in UFDs at least. But solving even one of problems is quite hard with ring theory alone, and we need to combine our ideas of polynomial rings, groups, and vector spaces to develop effective means of solving polynomials. This is known as *Galois theory*.

One nice property where finding the roots and the factorization coincide is when $f(x)$ splits in $R[x]$, and we usually construct these rings that allow $f(x)$ to split by *extending* existing rings R into a bigger ring S.

Definition 11.1 (Ring Extension)

A pair of rings R, S where R is a subring of S is called a **ring extension**, denoted $R \hookrightarrow S$, S/R or $R \subset S$. If we have multiple extensions $R \subset S \subset T$, this is called a **tower**.

We have already come across several examples of extensions.

Example 11.1 (Ring Extensions)

Listed.
1. $\mathbb{Z} \subset \mathbb{Q}$ is a ring extension.
2. $\mathbb{Z} \subset \mathbb{Z}[i]$ is a ring extension.
3. $\mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$ is a ring extension tower.

Now we present the motivation for working with extensions of fields that rings in general. Recall that a field is trivially a vector space, followed by the more general result.

Theorem 11.1 (Fields are a Vector Space over Subfields)

Let F be a subfield of K. Then K is a F-vector space.

Proof. A F-vector space has 0, addition, and multiplication by F. K indeed has 0, addition, and we can multiply any element of K by an element of F. The extra axioms follow but are too verbose to write a full proof.

Corollary 11.2

$\mathbb{R}$ is an infinite-dimensional vector space over $\mathbb{Q}$.

Proof. The fact that it is a vector space immediately follows from $\mathbb{Q} \hookrightarrow \mathbb{R}$. For dimensionality, the outline is to show that $\{\sqrt{p} \mid p \text{ prime } \}$ are linearly independent. This takes work to prove and won't do it.

Therefore, by constructing a subfield F of a field K, we can model K as a F-vector space (though it may be finite or infinite-dimensional). This additional structure warrants a name.

Definition 11.2 (Field Extension)

If $F \subset K$ are fields, then this is called a **field extension**. Its **degree** is the F-dimension of K, denoted

$$[K : F] \coloneqq \dim_F(K) \tag{525}$$

As we concatenate field extensions, sometimes called a *tower*, the dimensions behave nicely as well.

Theorem 11.3 (Tower Rule)

If $E \hookrightarrow F \hookrightarrow K$ are two field extensions, then $E \hookrightarrow K$ is a field extension with degree

$$[K : E] = [K : E][E : F] \tag{526}$$

Proof. Let $\alpha_1, \dots \alpha_m$ be a basis for E over F and $\beta_1, \dots, \beta_n$ be a basis for K over E. We claim that $\{\alpha_i \beta_j\}$ is a basis for K over F, with multiplication done in the field K. We check linear indepdence. Let $\beta \in K$ be arbitrary. Then by the E-basis, we have

$$\beta = \sum_{j=1}^{n} x_j \beta_j \tag{527}$$

But since $x_j \in E$, there are elements

$$x_j = \sum_{i=1}^{m} y_{ij} \alpha_i \tag{528}$$

and so combining we get

$$\beta = \sum_{j=1}^{n} \sum_{i=1}^{m} y_{ij} (\alpha_i \beta_j) \tag{529}$$

To prove linear independence, suppose $\beta = 0$. Then we have

$$0 = \sum_{j=1}^{n} \sum_{i=1}^{m} y_{ij} (\alpha_i \beta_j) = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} y_{ij} \alpha_i \right) \beta_j \tag{530}$$

Since $\beta_1, \dots, \beta_n$ are linearly indpendent, we must have $\sum_{i=1}^{m} y_{ij} \alpha_i = 0$ for all j. But since α_i's are linear independent, this means $y_{ij} = 0$ for all i, j.

In fact, it is a natural question to find the "minimal field" F such that $f(x)$ splits in $F[x]$, which is called the *splitting field* of $f(x)$. To do this, we must start with F and extend it to K such that K is a field. This is done through *adjoining* an element α to F, getting a larger set $F[\alpha]$. This guarantees both a ring and vector space structure but requires another condition (that α be a root of an irreducible polynomial in $F[x]$) to be a field. Once this is done, we have successfully created a field extension, and finding the splitting field is not too hard from here.

What's next? Now that we can construct the splitting field K/F, we want to try and find some *symmetries* between its roots. That is, if α, β are the roots of some $f(x) \in F[x]$, can we find a transformation group G such that knowing one root allows us to uncover the other roots? There is in fact such a structure out there, known as the *Galois group*—an automorphism group on the finite set of roots. This will be our roadmap in this section.

11.1 Ring Extensions Through Adjoining

We deal with the first problem of constructing ring extensions. We have already seen the idea of taking ring R and adjoining an element x to it to get a minimal ring containing both R (as a subring) and x. This was the polynomial ring. Now we can replace the indeterminate x with an element α from a larger ring S/R and do the exact same thing to get a ring extension of R.

Definition 11.3 (Adjoining Ring)

Given ring extension $R \hookrightarrow S$ with finite $\alpha = \{\alpha_1, \dots, \alpha_n\} \subset S$, the **ring R adjoined by α** is defined in the two equivalent ways.

1. $R[\alpha]$ is the minimal ring containing R and α.
2. We take the α_i's and map it through all polynomials in $R[x_1, \dots, x_n]$.

$$R[\alpha] = R[\alpha_1, \dots, \alpha_n] \coloneqq \{f(\alpha_1, \dots, \alpha_n) \in S \mid f \in R[x_1, \dots, x_n]\} \subset S \tag{531}$$

3. We can iteratively construct $R[\alpha]$.[a]

$$R \subset R[\alpha_1] \subset R[\alpha_1, \alpha_2] \subset \dots \subset R[\alpha_1, \dots, \alpha_n] \tag{532}$$

[a] Note that if $R \hookrightarrow S$ is a ring extension, then $R[\alpha]$ is a ring, and so it makes sense to write $R[\alpha][\beta] \subset S$. It had better be the case that this ring is consistent with equivalent constructions.

Proof. We show a few things.

1. *$R[\alpha]$ in (2) is a ring.* Given two elements $\phi, \gamma \in F[\alpha]$, there exists polynomials $f, g \in F[x]$ s.t. $\phi = f(\alpha), \gamma = g(\alpha)$. Since $F[x]$ is a ring, we see that

$$\phi + \gamma = f(\alpha) + g(\alpha) = (f+g)(\alpha) \tag{533}$$
$$\phi \cdot \gamma = f(\alpha) \cdot g(\alpha) = (fg)(\alpha) \tag{534}$$

   Furthermore, it is easy to check that 0 and 1 are the images of α through the 0 and 1 polynomials.
2. *(1) $\implies$ (2).* This is pretty obvious since R is a subring of (2) and α is just the image of α under $f(x) = x$.
3. *(2) $\iff$ (3).* By induction, it suffices to show that $R[\alpha, \beta] = R[\alpha][\beta] = R[\beta][\alpha]$.
4. *(3) $\implies$ (1).* By induction, it suffices to show that $R[\alpha_1]$ is the minimal field a

What allows us to make this inclusion proper is that the $\alpha \in K$, which does not necessarily have to be in F, *extends* this field a bit further, but since we can only map the one element α, it may not cover all of K. Most of the times, we will work with adjoining rings of univariate polynomial elements.

$$F[\alpha] \coloneqq \{f(\alpha) \in F \mid f \in F[x]\} \subset K \tag{535}$$

Let's go through some examples.

Example 11.2 (Radical Extensions of $\sqrt{2}$)

Let $F = \mathbb{Q}$ and $K = \mathbb{C}$. We claim $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$.

1. $\mathbb{Q}[\sqrt{2}] \subset \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$. $\mathbb{Q}[\sqrt{2}]$ are elements of the form

$$f(\sqrt{2}) = a_n(\sqrt{2})^n + a_{n-1}(\sqrt{2})^{n-1} + \dots + a_2(\sqrt{2})^2 + a_1\sqrt{2} + a_0 \tag{536}$$

   This can be written by collecting terms, of the form $a + b\sqrt{2}$.
2. $\mathbb{Q}[\sqrt{2}] \supset \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$. Given an element $a + b\sqrt{2}$, this is clearly in $\mathbb{Q}[\sqrt{2}]$ since it is the image of $\sqrt{2}$ under the polynomial $f(x) = a + bx$.

Given this, we may extrapolate this pattern and claim that $\mathbb{Q}[\sqrt{2} + \sqrt{3}]$ consists of all numbers of form $a + (\sqrt{2} + \sqrt{3})b$. However, this is *not* the case.

Example 11.3 (Radical Extensions of $\sqrt{2}+\sqrt{3}$)

Given any element $\beta \in \mathbb{Q}[\sqrt{2}+\sqrt{3}]$, it is by definition of the form

$$\beta = \sum_{k=0}^{n} a_k(\sqrt{2}+\sqrt{3})^k \tag{537}$$

Clearly $1, \sqrt{2}+\sqrt{3} \in \mathbb{Q}[\sqrt{2}+\sqrt{3}]$ by mapping $\sqrt{2}+\sqrt{3}$ through the polynomials $f(x)=1$ and $f(x)=$. However, we can see that $(\sqrt{2}+\sqrt{3})^2 = 5+\sqrt{6}$,[a] and so $\sqrt{6} \in \mathbb{Q}[\sqrt{2}+\sqrt{3}]$. Furthermore, we have $(\sqrt{2}+\sqrt{3})^3 = 11\sqrt{2}+9\sqrt{3}$, and so with the ring properties we can conclude that

$$\frac{1}{2}\big[(11\sqrt{2}+9\sqrt{3}) - 9(\sqrt{2}+\sqrt{3})\big] = \sqrt{2} \in \mathbb{Q}[\sqrt{2}+\sqrt{3}] \tag{538}$$

$$-\frac{1}{2}\big[(11\sqrt{2}+9\sqrt{3}) - 11(\sqrt{2}+\sqrt{3})\big] = \sqrt{3} \in \mathbb{Q}[\sqrt{2}+\sqrt{3}] \tag{539}$$

$$\tag{540}$$

If we go a bit further, we can show that

$$\mathbb{Q}[\sqrt{2}+\sqrt{3}] = \{a+b\sqrt{2}+c\sqrt{3}+d\sqrt{6} \mid a,b,c,d \in \mathbb{Q}\} \tag{541}$$

[a] where we use $\sqrt{6}$ as notation for $\sqrt{2}\cdot\sqrt{3}$

Example 11.4 (Showing Two Extensions are Equal)

We claim that $\mathbb{Q}[\sqrt{6},\sqrt{5}] = \mathbb{Q}[\sqrt{6}+\sqrt{5}]$.

1. $\mathbb{Q}[\sqrt{6},\sqrt{5}] \subset \mathbb{Q}[\sqrt{6}+\sqrt{5}]$ because the sum of $\sqrt{6}$ and $\sqrt{5}$ lies in $\mathbb{Q}[\sqrt{6},\sqrt{5}]$.
2. $\mathbb{Q}[\sqrt{6}+\sqrt{5}] \subset \mathbb{Q}[\sqrt{6},\sqrt{5}]$. We have

$$\frac{1}{\sqrt{5}+\sqrt{6}} = \frac{1}{\sqrt{5}+\sqrt{6}}\frac{\sqrt{5}-\sqrt{6}}{\sqrt{5}-\sqrt{6}} = \sqrt{6}-\sqrt{5} \in \mathbb{Q}[\sqrt{5}+\sqrt{6}] \tag{542}$$

which implies that

$$\sqrt{6} = \frac{1}{2}((\sqrt{6}+\sqrt{5}) + (\sqrt{6}-\sqrt{5})) \in \mathbb{Q}[\sqrt{5}+\sqrt{6}] \tag{543}$$

and so $\sqrt{5} = (\sqrt{6}+\sqrt{5}) - \sqrt{6} \in \mathbb{Q}[\sqrt{5}+\sqrt{6}]$.

So far, we've gotten used to modeling fields as vector spaces. In general, an adjoining ring $R[\alpha]$ is another ring containing R as a subring. Now if F is a field, then $F \subset F[\alpha]$ is a ring extension, and since we are working with the base field F, we can model $F[\alpha]$ as a vector space.

Theorem 11.4 (Adjoining Field is Finite-Dimensional Vector Space)

If $F \subset K$ is a field and $\alpha \in K$, then

1. $F[\alpha]$ is a *finite-dimensional* vector space over F.
2. If $f(x) = a_n x^n + \ldots + a_0$ is any polynomial with root α, then $\{1, \alpha, \ldots, \alpha^{n-1}\}$ spans $F[\alpha]$.[a]

[a] Note that this does not mean that it is a basis.

Proof. An element of $F[\alpha]$ is of the form

$$f(\alpha) = \sum_{k=0}^{n} a_k \alpha^k \tag{544}$$

for some $f \in F[x]$, and so it is immediate that $\{\alpha^k\}_{k \in \mathbb{N}_0}$ spans $F[\alpha]$. We claim that α^{n-1+i} is in S for all $i > 0$. By induction, if $i = 1$, then

$$\alpha^n = -\frac{1}{a_n}\big(a_{n-1}\alpha^{n-1} + \ldots + a_0\big) \tag{545}$$

which proves the claim. Now assume that $\alpha^n, \alpha^{n+1}, \ldots, \alpha^{n-1+i} \in \text{span}\{1, \ldots, \alpha^{n-1}\}$. Then

$$\alpha^i f(\alpha) = 0 \implies a_n\alpha^{n+i} + \alpha_{n-1}\alpha^{n+i-1} + \ldots + a_0\alpha^i = 0 \tag{546}$$

and so

$$\alpha^{n+i} = -\frac{1}{a_n}\big(a_{n-1}\alpha^{n+i-1} + \ldots + a_0\alpha^i\big) \tag{547}$$

which means that $\alpha^{n+i} \in \text{span}\{1, \ldots, \alpha^{n-1}\}$, completing the proof.

11.2 Field Extensions

This is nice since we have a vector space structure on $F[\alpha]$ unlike just a ring structure on $R[\alpha]$. What we should think is that $F[\alpha]$ ends up becoming both a ring and a vector space, but not yet a field. We would like to find conditions in which $F[\alpha]$ indeed does become a field, which at this point it is commonly denoted $F(\alpha)$ (rather than square brakcets to emphasize it is a field). It turns out that it will happen if any only if α is *algebraic*.

Definition 11.4 (Algebraic Numbers)

Given a field extension $F \subset K$, an element $\alpha \in K$ is **algebraic over** F if α is some root of $f(x) \in F[x]$.[a]

[a] By default we have $F = \mathbb{Q}$.

Example 11.5 (Algebraic Numbers)

We list a few examples.

1. $\sqrt{2} \in \mathbb{R}$ is algebraic over $\mathbb{Q}$ since it is a root of $x^2 - 2 \in \mathbb{Q}[x]$.
2. $i \in \mathbb{C}$ is algebraic over $\mathbb{R}$ since it is a root of $x^2 + 1 \in \mathbb{R}[x]$. It is also a root of $x^4 + 2x^2 + 1 \in \mathbb{R}[x]$.

Note that given an element $\alpha \in K$ that is algebraic over F, there may be multiple polynomials $f(x) \in F[x]$ that has α as a root. In fact, this set forms an ideal.

Lemma 11.5 (Polynomials Vanishing at α Forms an Ideal)

Let K/F be a field extension and fix $\alpha \in K$ that is algebraic over F. Then the set of polynomials $f(x) \in F[x]$ with root α forms an ideal over $F[x]$.

Proof. Let us denote the set as I. We prove the two properties of ideals.

1. Consider $f(x), g(x) \in I$. Then $(f+g)(\alpha) = f(\alpha) + g(\alpha) = 0 + 0 = 0 \implies (f+g)(x) \in I$.
2. Consider $h(x) \in F[x], f(x) \in I$. Then $(hf)(\alpha) = h(\alpha)f(\alpha) = h(\alpha)0 = 0 \implies (hf)(x) \in I$.

Since F is a field, $F[x]$ is a PID, and so it must be generated by some element. The scaling of the coefficients doesn't really matter (since we are working in a field so we can always divide the leading coefficient), so we can assume that it is monic. This is called the *minimal polynomial*.

Definition 11.5 (Minimal Polynomial)

Let $F \subset K$ be a field extension and $\alpha \in K$. The **minimal polynomial** of α is defined in the equivalent ways.

1. It is the monic polynomial of least degree among all polynomials in $F[x]$ having α as a root.
2. It is the generator of the ideal of all polynomials in $F[x]$ with root α.

We claim that it always exists and is unique.

Proof. Take the evaluation homomorphism $\text{ev}_\alpha : F[x] \to K$ and look at the ideal $\ker \text{ev}_\alpha$. Since $F[x]$ is a PID, call the generator $f(x)$ and normalize the leading coefficient to 1. We claim that this is the minimal polynomial. This proves existence.
TBD.

Example 11.6 (Minimal Polynomials)

Let $\mathbb{Q} \subset \mathbb{R}$ be a field extension.

1. The minimal polynomial for $\alpha = \sqrt{2} \in \mathbb{R}$ in $\mathbb{Q}[x]$ is
$$x^2 - 2 = (x + \sqrt{2})(x - \sqrt{2}) \tag{548}$$
2. The minimal polynomial for $\alpha = \sqrt{2} \in \mathbb{R}$ in $\mathbb{R}[x]$ is
$$x - \sqrt{2} \tag{549}$$
3. The minimal polynomial for $\alpha = \sqrt{2} + \sqrt{3}$ in $\mathbb{Q}[x]$ is
$$x^4 - 10x^2 + 1 = (x - (\sqrt{2} + \sqrt{3}))(x - (\sqrt{2} - \sqrt{3}))(x - (-\sqrt{2} + \sqrt{3}))(x - (-\sqrt{2} - \sqrt{3})) \tag{550}$$

Note that the minimal polynomial is also irreducible. Let's prove this.

Lemma 11.6 (Minimal Polynomial is Irreducible)

Let K/F be a field extension. The minimal polynomial $f(x) \in F[x]$ of $\alpha \in K$ is irreducible in $F[x]$.

Proof. Assume that $f(x)$ is reducible. Then we factor it into $f(x) = g(x)h(x)$, which implies that
$$0 = f(\alpha) = g(\alpha)h(\alpha) \implies g(\alpha) = 0 \text{ or } h(\alpha) = 0 \tag{551}$$
since F is a field and therefore does not contain zero divisors. Choosing both $g(x), h(x)$ to be of degree strictly lower than $f(x)$ would contradict the minimality requirement of $f(x)$, so $f(x)$ must be irreducible.

Since the minimal polynomial is irreducible, we can see that from Bezout's identity that the minimal polynomial generates a maximal ideal, and so the quotient ring is actually a field.

Theorem 11.7 (Quotient Ring of Minimal Polynomial is a Field)

Let F be a field and $f(x) \in F[x]$, then

$$f(x) \text{ is irreducible in } F[x] \iff \frac{F[x]}{\langle f(x) \rangle} \text{ is a field} \tag{552}$$

Proof. We first claim that $f(x)$ is irreducible if and only if $f(x)$ is a maximal ideal.

1. If $f(x)$ is irreducible, then consider the ideal $I = \langle f(x) \rangle$ and add some $g(x) \in F[x] \setminus I$. Since $F[x]$ is a PID, $f(x)$ is prime, and so $\gcd(f, g) = 1$. From Bezout's identity, there must exist $a(x), b(x) \in F[x]$ s.t. $a(x)f(x) + b(x)g(x) = 1$, which implies that any larger ideal than $\langle f(x) \rangle$ must be the entire $F[x]$.
2. If $f(x)$ is reducible, then let $f(x) = g(x)h(x)$. Then $\langle f(x) \rangle \subsetneq \langle g(x) \rangle \subsetneq F[x]$, and so $\langle f(x) \rangle$ is not maximal.

Then with 7.8, we know that I is a maximal ideal if and only if $F[x]/I$ is a field, and the proof is complete.

Now we went through all this trouble of determining sufficient conditions for quotient rings to be fields, when our original goal was on adjoining rings. It seemed like a long detour, but this finally pays off, as now we have both a necessary and sufficient condition to reach our original goal.

Corollary 11.8 (Conditions For Adjoining Ring to be a Field)

Let $F \hookrightarrow K$ be a field extension and $\alpha \in K$.

1. If α is algebraic over F—i.e. it is the root of some $g(x) \in F[x]$—then $F(\alpha) \subset K$ is a field.
2. The F-dimension of K $[F(\alpha) : F]$ is the degree of the minimal polynomial of α.[a]

[a] It is clear that if there exists *some* $f(x) \in F[x]$ that has root α, then it may not be the *unique* one. The dimension resides specifically in the unique minimal polynomial.

Proof. Let $f(x)$ be the minimal polynomial of α of degree n, and we know that $g(x) \in \langle f(x) \rangle \implies g(x) = h(x)f(x)$ for some $h(x) \in F[x]$. Take the surjective ring homomorphism $\text{ev}_\alpha : F[x] \to F[\alpha]$. We know from 11.5 that $\ker(\text{ev}_\alpha) = \langle f(x) \rangle$, and by the first isomorphism theorem of rings, we have

$$\frac{F[x]}{\langle f(x) \rangle} = \frac{F[x]}{\ker(\text{ev}_\alpha)} \simeq F[\alpha] \tag{553}$$

which proves the first claim. For dimension, we know that $\{1, \dots, x^{n-1}\}$ is a basis.

Proof. This alternative proof is constructive in that it actually shows how to compute multiplicative inverses.

It is clear that $F[\alpha]$ is a commutative ring since F is a field. So it remains to show that every nonzero element of $\beta \in F[\alpha]$ is a unit. By definition $\beta = p(\alpha)$ for some polynomial $p \in F[x]$. Factor $f \in F[x]$ as the product of irreducible polynomials. Then α must be a root of one of those irreducible factors, say $g(x)$. Note that $g(x) \nmid p(x)$ since $p(\alpha) \neq 0$. Since g is irreducible, we know that $\gcd(g, p) = 1$ and so $\exists s, t \in F[x]$ s.t.

$$1 = sp + tg \implies 1 = s(\alpha)p(\alpha) + t(\alpha)g(\alpha) = s(\alpha)p(\alpha) \tag{554}$$

Therefore we have found a multiplicative inverse $s = p^{-1} \in F[\alpha]$.

Proof. We can also prove it using the vector space structure, though this is for finite-dimensional vector

spaces. Treating $F[\alpha]$as a finite-dimensional vector space over F, let us define the F-linear function[a]

$$m_b : F[\alpha] \to F[\alpha], \qquad m_b(\beta) = b\beta \tag{555}$$

Since $F[\alpha] \subset K$, $F[\alpha]$ is an integral domain. Thus $\nexists \beta \in F[\alpha] \setminus \{0\}$ s.t. $b\beta = 0$. This means that the kernel of m_b is 0, and so m_b is injective. By the rank-nullity theorem, it is bijective, and so there exists a $\beta \in F[\alpha]$ s.t. $b\beta = 1 \implies b$ is a unit.

[a] linearity is easy to check

Since $F[\alpha]$ is the smallest ring containing both F and α, it immediately follows that it is the smallest *field* containing F and α. Therefore, we have unified the two constructions of adjoining fields and quotient rings.

Example 11.7 (Some Adjoining Fields)

Here are some examples.

1. Given any algebraic number $\alpha \in \mathbb{C}$, by definition there exists a $f(x) \in \mathbb{Q}[x]$ with roots α, and so $\mathbb{Q}(\alpha) \subset \mathbb{C}$.[a]
2. $\mathbb{Q}(\sqrt{3}i)$ is a field of dimension 2 since $\sqrt{3}i$ is a root of the polynomial $f(x) = x^2 + 3$.
3. However, $\mathbb{Q}[\pi]$ is not a field since π is a *transcendental number*. However we will not prove it now.

[a] In fact it was historically so common to work solely in subfields of $\mathbb{C}$ that an *algebraic number* meant by default algebraic with respec to the field extension $\mathbb{C}/F$.

Note that we now have the field extension tower $F \subset F[\alpha] \subset K$. By the tower property, $[K : F] = [F : F(\alpha)][F(\alpha) : F]$, and so $\deg f(x) = [F : F(\alpha)]$ must divide $[K : F]$.

Now recall quotient rings, which do not necessarily preserve the properties of the original ring. That is, if F is a field, then F/I may not be a field. Using the fundamental ring homomorphism theorem, we can precisely correlate certain quotient maps with adjoining fields. Recall that given a field extension $F \subset K$, the evaluation function $\text{ev}_\alpha : F[x] \to K$ defined $f(x) \mapsto f(\alpha)$ is a homomorphism.

Example 11.8 (Simple Quotient Rings as Field Adjoined with 1 Variable)

Consider the following.

1. Since $x^2 + 1 \in \mathbb{Z}_7[x]$ is irreducible, $\mathbb{Z}_7[x]/\langle x^2 + 1\rangle$ is a field.
2. Since $x^2 + 1 \in \mathbb{R}[x]$ is irreducible of degree 2, so the quotient ring is a field. Furthermore, $i \in \mathbb{C}$ is a root of the degree 2 minimal polynomial, so we have the isomorphism

$$\frac{\mathbb{R}[x]}{\ker(\text{ev}_i)} = \frac{\mathbb{R}[x]}{\langle x^2 + 1\rangle} \simeq \mathbb{R}[i] = \mathbb{C} \tag{556}$$

induced by the evaluation map

$$\phi : \frac{\mathbb{R}[x]}{\langle x^2 + 1\rangle} \to \mathbb{C}, \qquad \phi\big(f(x) \pmod{\langle x^2 + 1\rangle}\big) = f(i) \tag{557}$$

Therefore $\mathbb{R}(i)$ has $\mathbb{R}$-dimension 2, which makes sense since we view $\mathbb{C}$ as being isormophic to $\mathbb{R}^2$.

3. The evaluation map

$$\text{ev}_{\sqrt{2}} : \mathbb{Q}[x] \mapsto \mathbb{Q}[\sqrt{2}], \qquad \text{ev}_{\sqrt{2}}(f) = f(\sqrt{2}) \tag{558}$$

is a homomorphism. Furthermore, it has a kernel $\langle x^2 - 2\rangle$ since $(x^2 - 2)$ is an irreducible polynomial in $\mathbb{Q}[x]$ containing the root $\sqrt{2}$. Therefore by the fundamental ring homomorphism theorem we have

$$\frac{\mathbb{Q}[x]}{\langle x^2 - 2\rangle} \simeq \mathbb{Q}[\sqrt{2}] \tag{559}$$

Example 11.9 (Extensions of $\sqrt{2}$ and i)

We claim that

$$\mathbb{Q}[\sqrt{2}, i] = \{a + b\sqrt{2} + ci + d(\sqrt{2}i) \mid a, b, c, d \in \mathbb{Q}\} \tag{560}$$

From the previous example, we know that $\mathbb{Q}[\sqrt{2}]$ are all numbers of the form $a + b\sqrt{2}$. Now we take $i \in \mathbb{C}$ and map it through all polynomials with coefficients in $\mathbb{Z}[\sqrt{2}]$, which will be of form

$$f(i) = (a_n + b_n\sqrt{2})i^n + (a_{n-1} + b_{n-1}\sqrt{2})i^{n-1} + \ldots + (a_2 + b_2\sqrt{2})i^2 + (a_1 + b_1\sqrt{2})i + (a_0 + b_0\sqrt{2}) \tag{561}$$

However, we can see that since $i^2 = -1$, we only need to consider up to degree 1 polynomials of form

$$(a + b\sqrt{2}) + (c + d\sqrt{2})i \tag{562}$$

which is clearly of the desired form. For the other way around, this is trivial since we can construct a linear polynomial as before.

Example 11.10

We claim $\mathbb{Q}[\sqrt{3} + i] = \mathbb{Q}[\sqrt{3}, i]$.

1. $\mathbb{Q}[\sqrt{3} + i] \subset \mathbb{Q}[\sqrt{3}, i]$
2. $\mathbb{Q}[\sqrt{3} + i] \supset \mathbb{Q}[\sqrt{3}, i]$. Note that

$$(\sqrt{3} + i)^3 = 8i \implies i \in \mathbb{Q}[\sqrt{3} + i] \tag{563}$$

$$\implies (\sqrt{3} + i) - i = \sqrt{3} \in \mathbb{Q}[\sqrt{3} + i] \tag{564}$$

Therefore, $\mathbb{Q}[\sqrt{3} + i]$ contains the elements $1, \sqrt{3}, i$, which form the basis of $\mathbb{Q}[\sqrt{3}, i]$.

Example 11.11 (Extensions of $\sqrt{3}i$ and $\sqrt{3}, i$)

We claim that $\mathbb{Q}[\sqrt{3}i] \subsetneq \mathbb{Q}[\sqrt{3}, i]$.

1. We can see that $\{1, \sqrt{3}i\}$ span $\mathbb{Q}[\sqrt{3}i]$ as a $\mathbb{Q}$-vector space. Therefore,

$$\sqrt{3}, i \in \mathbb{Q}[\sqrt{3}, i] \implies \sqrt{3}i \in \mathbb{Q}[\sqrt{3}, i] \tag{565}$$

implies that $\mathbb{Q}[\sqrt{3}i] \subset \mathbb{Q}[\sqrt{3}, i]$.

2. To prove proper inclusion, we claim that $i \notin \mathbb{Q}[\sqrt{3}i]$. Assuming that it can, we represent it in the basis $i = b_0 + b_1\sqrt{3}i$, and so

$$-1 = (b_0 + b_1\sqrt{3}i)^2 = (b_0^2 - 3b_1^2) + 2b_0b_1\sqrt{3}i \tag{566}$$

Therefore we must have $2b_0b_1\sqrt{3} = 0 \implies b_0$ or b_1 should be 0. If $b_0 = 0$, then $b_0^2 - 3b_1^2 = -3b_1^2 \implies b_1^2 = 1/3$, which is not possible since $b_1^2 \in \mathbb{Q}$. If $b_1 = 0$, then $b_0 - 3b_1^2 = b_0^2 > 0$, and so it cannot be -1.

The most significant property is that a field contains inverses. So how do we compute such an inverse?

Example 11.12 (Computing Inverses in Field Extensions)

We can do two problems.

1. Given $\beta = p(\alpha) = \alpha^2 + \alpha - 1 \in \mathbb{Q}[\alpha]$, where α is a root of $f(\alpha) = \alpha^3 + \alpha + 1$, we first know that β must have a multiplicative inverse since $\mathbb{Q}[\alpha]$ is a field. Applying the Euclidean algorithm, we

have

$$1 = \frac{1}{3}\big\{(x+1)f(x) - (x^2+2)p(x)\big\} = -\frac{1}{3}(\alpha^2+2)p(\alpha) \tag{567}$$

and so $\beta^{-1} = (\alpha^2+\alpha-1)^{-1} = -\frac{1}{3}(\alpha^2+2)$. We can check that

$$-\frac{1}{3}(\alpha^2+2)(\alpha^2+\alpha-1) = -\frac{1}{3}(\alpha^4+\alpha^3+\alpha^2+2\alpha-2) \tag{568}$$

$$= -\frac{1}{3}(\alpha^3+\alpha-2) \tag{569}$$

$$= -\frac{1}{3}(-3) = 1 \tag{570}$$

2. Given $f(x) = x^3+3 \in \mathbb{Q}[x]$, we know (from Eisenstein) that it is irreducible. Therefore we know that $\mathbb{Q}[x]/\langle f(x)\rangle$ is a field. Furthermore, it contains a root α, and so

$$\frac{F[x]}{\langle x^3+x\rangle} \simeq F[\alpha] \tag{571}$$

Now take an arbitrary element of the field, say $f(\alpha) = \alpha^2 - 1$. Then it must have an inverse call it $g(\alpha)$ that satisfies

$$1 = f(\alpha)g(\alpha) \text{ in } \frac{\mathbb{Q}[x]}{\langle x^3+3\rangle} \tag{572}$$

But by working in the original ring $\mathbb{Q}[x]$ and not the quotient, this would look something like.

$$1 = f(x)g(x) + k(x)(x^3+1) \text{ in } \mathbb{Q}[x] \tag{573}$$

for some $k(\alpha)$. Note that we are guaranteed to get this form since $f(x)$ is irreducible and so $\gcd(f,g) = 1$. So we use Euclidean division.

$$x^3+3 = (x)(x^2-1) + (x+3) \tag{574}$$

$$x^2-1 = (x-3)(x+3) + 8 \tag{575}$$

Therefore by substituting we have

$$8 = (x-3)(x^3+3) - (x^2-2x-1)(x^2-1) \tag{576}$$

and so by taking $\pmod{x^3+3}$, we have

$$1 = -\frac{1}{8}(x^2-3x-1)(x^2-1) \implies \frac{1}{\alpha^2-1} = -\frac{1}{8}(\alpha^2-3\alpha-1) \tag{577}$$

Example 11.13 (Irreducible Quotient Rings)

On the other hand, if we consider the ring $\mathbb{Z}_5[x]/\langle x^2-2\rangle$, then x^2-2 is irreducible in $\mathbb{Z}_5[x]$ (just plug in $0,1,2,3,4$), and so since $\mathbb{Z}_5$ is a field, the quotient ring is a field. Then we have

$$\frac{\mathbb{Z}_5[x]}{\langle x^2-2\rangle} \simeq \mathbb{Z}_5[\sqrt{2}] = \{a+b\sqrt{2} \mid a,b \in \mathbb{Z}_5\} \tag{578}$$

Remember that we should not be careless and just mindlessly assume that every quotient ring is a field. This is only when the ideal is generated by a minimal polynomial. Here is an example where the ideal is reducible, but fortunately using the chinese remainder theorem, we can first reduce it as a direct product of quotient rings, and then apply our theorem to represent them as a field structure.

Example 11.14 (Quotient Ring as Product Rings with Chinese Remainder Theorem)

Consider the following ring $\mathbb{Z}_5[x]/\langle x^2+1\rangle$. Then x^2+1 is reducible since $x^2+1=(x+2)(x+3)$, which implies $\langle x^2+1\rangle = \langle x+2\rangle \cap \langle x+3\rangle$, and so by the Chinese remainder theorem, we have

$$\frac{\mathbb{Z}_5[x]}{\langle x^2+1\rangle} = \frac{\mathbb{Z}_5[x]}{\langle x+2\rangle \cap \langle x+3\rangle} = \frac{\mathbb{Z}_5[x]}{\langle x+2\rangle} \times \frac{\mathbb{Z}_5[x]}{\langle x+3\rangle} \simeq \mathbb{Z}_5 \times \mathbb{Z}_5 \tag{579}$$

11.3 Splitting Fields

Great, so we have established conditions for which a field adjoined with an element becomes another field. All that is left to do is to find a splitting field.

Remember that by previously establishing that $\mathbb{C}$ is algebraically closed, this gives us a "safe space" to work in, in the sense that if we take any subfield $F \subset \mathbb{C}$ and find a polynomial $f(x) \in F[x]$, we are *guaranteed* to find a linear factorization of f in $\mathbb{C}[x]$. Therefore, if K is algebraically closed and $F \subset K$ is a field extension, $f(x) \in F[x]$ is guaranteed to *split* completely into linear factors. This is true for *all* $f(x) \in F[x]$, but now if we *fix* $f(x) \in F[x]$, perhaps we don't need the entire field K to split $f(x)$. Maybe we can work in a slightly larger field E—such that $F \subset E \subset K$—where $f(x)$ splits in E. This process of finding such a minimal field is important to understand the behavior of roots of such polynomials.

Definition 11.6 (Splitting Field)

Given a field extension $F \subset K$ and a polynomial $f \in F[x]$,

1. f **splits** in K if f can be written as the product of linear polynomials in $K[x]$.
2. If f splits in K and there exists no field E s.t. $F \subsetneq E \subsetneq K$, then K is called a **splitting field** of f.[a]

We claim that a splitting field always exists for any $f(x) \in F[x]$.

[a] i.e. the splitting field is the "smallest" field that splits f.

Proof. Let us decompose $f(x)$ into irreducible factors over F: $f(x) = f_1(x)f_2(x)\cdots f_k(x)$. If all of them are linear, then F is the splitting field for $f(x)$. Otherwise, we can assume that $f_1(x)$ is not linear. Then we can consider the field $K_1 = F[x]/\langle f_1(x)\rangle$, where $f_1(x)$ has a root α. Then $f_1(x) = (x-\alpha)g_1(x)$, and $f(x)$ has at least one (but maybe more) linear factor over K_1. If all irreducible factors over K_1 are linear, stop, otherwise there is an irreducible factor of degree at least 2, and we can repeat the procedure and add its root. Since a polynomial of degree n has at most n roots, the process will eventually stop and all factors will be linear in some extension of F.

It also turns out to be unique but we will prove this at the end of the section.

Example 11.15 (Don't Need Necessarily Complex Numbers to Split)

Consider the following subfields of $\mathbb{C}$ and observe that they are enough to split a given polynomial.

1. Let $f(x) = x^2 - 1$. If $f(x) \in \mathbb{R}[x]$, it does split in $\mathbb{R}$. In fact, even if we consider it as an element of $\mathbb{Z}_2[x]$, it still splits into $(x+1)(x-1)$.
2. Let $f(x) = x^2 - 2$. If $f(x) \in \mathbb{Q}[x]$, it doesn't split in $\mathbb{Q}$ since the roots $\pm\sqrt{2} \notin \mathbb{Q}$, but $\pm\sqrt{2}$ are real numbers, so $f(x)$ does in fact split in $\mathbb{R}$ since it splits into $(x+\sqrt{2})(x-\sqrt{2})$. However, maybe it is not the (smallest) splitting field.
3. Let $f(x) = x^2 + 1$. We can see that if we consider it as an element of $\mathbb{Q}[x]$ or $\mathbb{R}[x]$, neither fields split $f(x)$ since $\pm i$ are its roots and therefore are contained in the coefficients of its linear factors. We know that it definitely splits in $\mathbb{C}$, but can we find a smaller field that splits $f(x)$? Perhaps.

So how does one find a splitting field? Note that in the example above, we have found that there were some roots α of certain polynomials $f(x) \in F[x]$ are not contained in F. Therefore, what we want to do is find the smallest field F containing both F and α (plus any other α's). This is precisely the adjoining field $F[\alpha]$, which guarantees unique factorization since $F[\alpha]$ is a Euclidean domain.

Lemma 11.9 (Square-Free Extensions as Splitting Field)

If a is not a perfect square in F then $F(\sqrt{a})$ is the splitting field of $f(x) = x^2 - a$.

Proof. TBD

We first provide some straightforward examples of computing splitting fields.

Example 11.16 (Straightforward Computation of Splitting Fields)

For some polynomials, finding their roots is trivial.

1. Let $f(x) = (x^2 - 2) \in \mathbb{Q}[x]$. Then the splitting field is $\mathbb{Q}(\sqrt{2})$.
2. Let $f(x) = (x^2 - 2)(x^2 - 3) \in \mathbb{Q}[x]$. Then the splitting field is $\mathbb{Q}(\sqrt{2}, \sqrt{3})$.

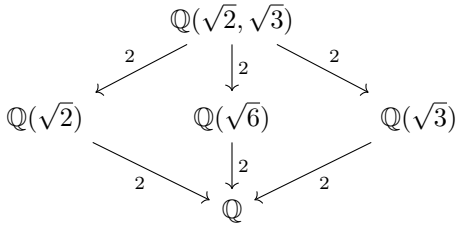


Figure 16: Diagram of known subfields of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$.

Note that

Example 11.17 (Slightly Harder Computation of Splitting Fields)

Note that often, the splitting field is "smaller" that one might suspect.

1. Let $f(x) = x^2 + 2x + 2 \in \mathbb{Q}[x]$. Then the roots of $f(x)$ are $-1 \pm i$, so
$$f(x) = (x - (-1 + i))(x - (-1 - i)) \tag{580}$$
and we can show that $\mathbb{Q}[-1 - i, -1 + i] = \mathbb{Q}[i]$ is the splitting field of f. It has dimension 2 since $f(x)$ is a 2nd degree polynomial that is minimal.
2. Let $f(x) = x^2 - 2x - 1 \in \mathbb{Q}[x]$. The roots are $1 \pm \sqrt{2}$, and so
$$f(x) = (x - (1 + \sqrt{2}))(x - (1 - \sqrt{2})) \tag{581}$$
and so $\mathbb{Q}[\sqrt{2}]$ is the splitting field of f. Since $f(x)$ is a 2nd degree polynomial that is minimal.
3. Let $f(x) = x^6 - 1 \in \mathbb{Q}[x]$. We can factor
$$f(x) = (x - 1)(x + 1)(x^2 + x + 1)(x^2 - x + 1) \tag{582}$$
and the non-rational roots are $\frac{\pm 1 \pm \sqrt{3}i}{2}$. Thus the splitting field of f is $\mathbb{Q}[\sqrt{3}i]$. The dimension is not 6 because $f(x)$ is reducible over $\mathbb{Q}$. It is 2. It is easier to look at this not as a quotient ring, but use the first theorem to find a minimal polynomial with root $\sqrt{3}i$. Indeed, such a polynomial is $x^2 + 3$, which has degree 2.

4. Let $f(x) = x^4 - 2 \in \mathbb{Q}[x]$, which has roots $\{\sqrt[4]{2}, \sqrt[4]{2}e^{\frac{2\pi i}{4}}, \sqrt[4]{2}e^{\frac{4\pi i}{4}}, \sqrt[4]{2}e^{\frac{6\pi i}{4}}\}$ and thus the splitting field is
$$\mathbb{Q}\big(\sqrt[4]{2}, \sqrt[4]{2}e^{\frac{2\pi i}{4}}, \sqrt[4]{2}e^{\frac{4\pi i}{4}}, \sqrt[4]{2}e^{\frac{6\pi i}{4}}\big) = \mathbb{Q}(\sqrt[4]{2}, e^{\frac{2\pi i}{4}}) \tag{583}$$
The inclusion $\subset$ is easy to prove, and to prove $\supset$, we can see that since we are working in a field,
$$e^{2\pi i/4} = \frac{\sqrt[4]{2}e^{2\pi i/4}}{\sqrt[4]{2}} \in \mathbb{Q}\big(\sqrt[4]{2}, \sqrt[4]{2}e^{\frac{2\pi i}{4}}, \sqrt[4]{2}e^{\frac{4\pi i}{4}}, \sqrt[4]{2}e^{\frac{6\pi i}{4}}\big) \tag{584}$$
which implies that $\sqrt[4]{2} \in \mathbb{Q}\big(\sqrt[4]{2}, \sqrt[4]{2}e^{\frac{2\pi i}{4}}, \sqrt[4]{2}e^{\frac{4\pi i}{4}}, \sqrt[4]{2}e^{\frac{6\pi i}{4}}\big)$. $f(x)$ is certainly minimal in $\mathbb{Q}[x]$ containing all roots, so the dimension of the field is 4. It is spanned by the powers of $\sqrt[4]{2}$.

Theorem 11.10 (Bounds on Degree of Splitting Field)

A splitting field of a polynomial $f(x) \in F[x]$ over F is of degree at most $n!$ over F.

Proof. The polynomial can have at most n roots, call them $\alpha_1, \dots, \alpha_n$.
1. Then, assume that F is irreducible and so $[F[\alpha_1] : F] = n$.
2. Now factor $f(x) = (x - \alpha_1)g(x) \in F[\alpha_1][x]$. If $g(x)$—which has degree $n-1$—is irreducible, $[F[\alpha_1, \alpha_2] : F[\alpha_1]] = n - 1$.

We will get the maximal degree at each step if the factored polynomial is irreducible. Then using the tower property, we have
$$[F[\alpha_1, \dots, \alpha_n] : F] = \prod_{i=0}^{n} [F[\alpha_1, \dots, \alpha_{i+1}] : F[\alpha_1, \dots, \alpha_i]] = n! \tag{585}$$

Example 11.18 (Splitting Field of $x^p - 1$)

Let p be prime, consider the polynomial $f(x) = x^p - 1$ over $\mathbb{Q}$. Let $\omega = e^{2\pi i/p}$, then ω is a root of $f(x)$ and we have
$$x^p - 1 = (x-1)(x-\omega)(x-\omega^2)\cdots(x-\omega^{p-1}), \tag{586}$$
so all roots of $f(x)$ belong to the extension $\mathbb{Q}(\omega)$. Since the minimal polynomial for ω equals $x^{p-1} + \dots + 1$, we have $[\mathbb{Q}(\omega) : \mathbb{Q}] = p - 1$.

Example 11.19 (Splitting Field of $x^p - 2$, p Prime)

Consider $f(x) = x^p - 2 \in \mathbb{Q}[x]$. If α is a root, i.e. $\alpha^p = 2$, then $(\zeta\alpha)^p = 2$, where ζ is any pth root of unity. Hence the p complex roots of $f(x)$ are $\zeta\sqrt[p]{2}$ for the p roots of unity ζ. It then follows that the splitting field is $\mathbb{Q}(\sqrt[p]{2}, \zeta)$. Now to compute the degree of this field extension, note that
$$[\mathbb{Q}(\sqrt[p]{2}, \zeta) : \mathbb{Q}] = [\mathbb{Q}(\sqrt[p]{2}, \zeta) : \mathbb{Q}(\sqrt[p]{2})][\mathbb{Q}(\sqrt[p]{2}) : \mathbb{Q}] = (p-1)p \tag{587}$$
which is true since it is easy to show that $[\mathbb{Q}[\sqrt[p]{2}] : \mathbb{Q}] = p$, and $x^2 - 2 = (x - \sqrt[p]{2})(x^{p-1} + \dots + 1) \in \mathbb{Q}[\sqrt[p]{2}]$, where $x^{p-1} + \dots + 1$ turns out to be irreducible—and hence the degree is $n - 1$. We could have done this the other way by decomposing
$$[\mathbb{Q}(\sqrt[p]{2}, \zeta) : \mathbb{Q}] = [\mathbb{Q}(\sqrt[p]{2}, \zeta) : \mathbb{Q}(\zeta)][\mathbb{Q}(\zeta) : \mathbb{Q}] = (p-1)p \tag{588}$$
which would yield the same result as proven in 11.3.

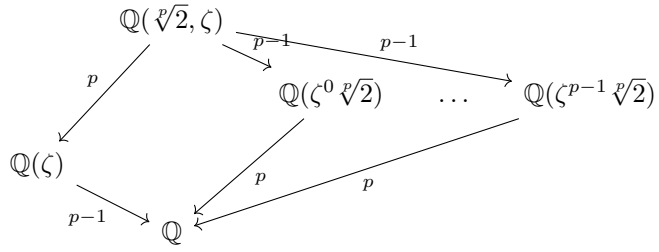


Figure 17: Diagram of known subfields of $\mathbb{Q}(\sqrt[p]{2}, \zeta)$.

Note that the order in which we added the adjoined the elements to the base field did not matter in the example above. We now formalize this.

Theorem 11.11 (Extending Isomorphisms of Fields)

Let $\phi : F \to F'$ be an isomorphism of fields.
1. Let $f(x) \in F[x]$ be an irreducible polynomial with root α in some field extension $K \supset F$.
2. Let $f'(x) = \phi(f(x)) \in F'[x]$ and let β be a root of $f'(x)$ in some field extension $K' \supset F'$.

Then there is a unique isomorphism $\bar{\phi} : F[\alpha] \to F'[\alpha']$ that is an extension of ϕ (i.e. behaves the same under F) and carries α to α.

$$\begin{array}{ccc} K & \xrightarrow{\bar{\phi}} & K' \\ \downarrow & & \downarrow \\ F & \xrightarrow{\phi} & F' \end{array}$$

Figure 18: Commutative diagram.

Proof. We use induction on the degree n of $f(x)$. Recall that a field (i.e. a ring) isomorphism $\phi : F \to F'$ induces a ring isomorphism $\tilde{\phi} : F[x] \to F'[x]$. jSo, if $f(x)$ and $f'(x)$ correspond to one another under this isomorphism, then the irreducible factors of $f(x) \in F[x]$ correspond to the irreducible factors of $f'(x) \in F'[x]$.

If $f(x)$ has all its roots in F then $f(x)$ splits completely in $F[x]$ and $f'(x)$ splits completely in $F'[x]$—with its linear factors being the images of the linear factors for $f(x)$. Hence $F = K$ and $F' = K'$, and in this case we may take $\bar{\phi} = \phi$. This shows the result is true for $n = 1$ and in the case where all the irreducible factors of $f(x)$ have degree 1.

Assume now by induction that the theorem holds for any field field F, isomorphism ϕ, and polynomial $f(x) \in F[x]$ of degree $< n$. Let $p(x)$ be an irreducible factor of $f(x) \in F[x]$ of degree at least 2 and let $p'(x)$ be the corresponding irreducible factor of $f'(x) \in F'[x]$. If $\alpha \in K$ is a root of $p(x)$ and $\beta \in K'$ is a root of $p'(x)$, then we can extend ϕ to an isomorphism $\phi' : F(\alpha) \to F'(\beta)$.

$$\begin{array}{ccc} F(\alpha) & \xrightarrow{\phi'} & F'(\beta) \\ \downarrow & & \downarrow \\ F & \xrightarrow{\phi} & F' \end{array}$$

Let $F_1 = F(\alpha), F_1' = F'(\beta)$, so that we have the isomorphism $\phi' : F_1 \to F_1'$. We have $f(x) = (x-\alpha)f_1(x)$ over F_1 with $\deg f(x) = n-1$ and $f'(x) = (x-\beta)f_1'(x)$. K is a splitting field for $f_1(x)$ over F_1 since: all the roots of $f_1(x)$ are in K and if they were contained in any smaller extension L containing F_1, then, since F_1 contains α, L would also contain all the roots of $f(x)$, which would contradict the minimality

of K as the splitting field of $f(x)$ over F. Similarly, K' is a splitting field for $f_1'(x)$ over F_1'. Since the degrees of $f_1(x)$ and $f_1'(x)$ are less than n, by induction there exists a map $\phi : K \to K'$ extending the isomorphism $\sigma' : F_1 \to F_1'$. This gives the extended diagram.

$$\begin{array}{ccc} K & \xrightarrow{\phi'} & K' \\ \downarrow & & \downarrow \\ F_1 & \xrightarrow{\phi'} & F_1' \\ \downarrow & & \downarrow \\ F & \xrightarrow{\phi} & F' \end{array}$$

Corollary 11.12 (Splitting Field is Unique)

The splitting field of $f(x) \in F[x]$ is unique up to isomorphism.

Proof. Take ϕ to be the identity mapping from F to itself, and K, K' to be two splitting fields for $f(x) = f'(x)$.

So far, we have fixed one polynomial and studied the unique splitting field of $f(x)$. Now what if we unfix $f(x)$?

Definition 11.7 (Algebraic Closure)

The field $\overline{F}$ is called an **algebraic closure** of F if $\overline{F}$ is algebraic over F and if every polynomial $f(x) \in F[x]$ splits completely over $\overline{F}$. That is, $\overline{F}$ contains all the elements algebraic over F.

Lemma 11.13 (Algebraic Closures are Algebraically Closed)

Let $\overline{F}$ be an algeraic closure of F, then $\overline{F}$ is algebraically closed.

Proof.

Theorem 11.14

For any field F there exists an algebraically closed field K containing F.

Proof.

In integral domains, by taking the field of fractions we can get a nice set of formulas often introduced in high-school math competitions.

Theorem 11.15 (Viete's Formulas)

Let $R[x]$ be an integral domain $f(x) \in R[x]$, and F be the field of fractions of R. If $f(x)$ splits in an

algebraically closed field extension K^a then

$$f(x) = a_0 \prod_{i=1}^{n} (x - \alpha_i) \tag{589}$$

for some $\alpha_1, \dots, \alpha_n \in K$, then the coefficients of f can be presented with the formulas

$$\sum_{i=1}^{n} \alpha_i = -\frac{a_1}{a_0} \tag{590}$$

$$\sum_{i_1 < i_2} \alpha_{i_1} \alpha_{i_2} = \frac{a_2}{a_0} \tag{591}$$

$$\sum_{i_1 < \dots < i_k} \prod_{j=1}^{k} \alpha_{i_j} = (-1)^k \frac{a_k}{a_0} \tag{592}$$

$$\alpha_1 \alpha_2 \alpha_3 ... \alpha_n = (-1)^n \frac{a_n}{a_0} \tag{593}$$

[a] Normally we take $R = \mathbb{Z}$, its field of fractions to be $\mathbb{Q}$, and its algebraically closed extension to be K.

Proof.

11.4 Finite Fields and Separability of Extensions

One property of polynomials that we have defined—yet have not studied much—was the multiplicity of its roots. Let's just introduce a quick definition and then provide a nice theorem to check muliplicity of roots.

Definition 11.8 (Separability of Polynoials)

A polynomial $f(x) \in F$ is **separable** if it has no repeated roots, i.e. no root of multiplicity greater than 1. A polynomial which is not separable is called inseparable.

Now we introduce the criterion to check separability. We introduce the derivative, which does coincide with the definition seen in analysis, but note that this is purely algebraic and should not be seen to have any connection with derivatives in analysis.

Definition 11.9 (Derivative)

The **derivative** of a polynomial $f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 \in F[x]$ is defined as an operator $D_x : F[x] \to F[x]$

$$D_x f(x) = f'(x) \coloneqq n a_n x^{n-1} + (n-1) a_{n-1} x^{n-2} + \dots + 2a_2 x + a_1 \in F[x] \tag{594}$$

Lemma 11.16 (Properties of the Polynomial Derivative)

$D_x : F[x] \to F[x]$ satisfies the properties.

1. *Linearity.*

$$D_x(cf)(x) = c D_x f(x) \tag{595}$$

$$D_x(f+g)(x) = D_x f(x) + D_x g(x) \tag{596}$$

2. *Product Rule.*
$$D_x(fg)(x) = f(x)\big(D_x g(x)\big) + \big(D_x f(x)\big)g(x) \tag{597}$$

Proof. Trivial through straightforward computation.

Theorem 11.17 (Conditions for Separability)

Let $f(x) \in F[x]$.
1. $f(x)$ has multiple roots α if and only if α is also a root of $D_x f(x)$, i.e. $f(x)$ and $D_x f(x)$ are both divisible by the minimal polynomial for α.
2. $f(x)$ is separable if and only if it is relatively prime to $D_x f(x)$, i.e. $\gcd\big(f(x), D_x f(x)\big) = 1$.

Proof. We prove bidirectionally.
1. ($\rightarrow$). Suppose that α is a multiple root of $f(x)$. Then over a splitting field,
$$f(x) = (x-\alpha)^n g(x) \tag{598}$$
for some $n \geq 2$ and some polynomial $g(x)$. Taking derivatives we get
$$D_x f(x) = n(x-\alpha)^{n-1} g(x) + (x-\alpha)^n D_x g(x) \tag{599}$$
which implies that α is a root of $(D_x f)(x)$.
2. ($\leftarrow$). Conversely, suppose that α is a root of both $f(x)$ and $D_x f(x)$. Then we write $f(x) = (x-\alpha)h(x)$ for some polynomial $h(x)$ and taking the derivative, we get
$$D_x f(x) = h(x) + (x-\alpha)D_x h(x) \implies 0 = D_x f(\alpha) = h(\alpha) \tag{600}$$
and so α is a root of h. Hence $h(x) = (x-\alpha)h_1(x)$ for some polynomial $h_1(x)$, and so $f(x) = (x-a)^2 h_1(x)$.

Corollary 11.18 (Separability over Field of Characteristic 0)

Every irreducible polynomial over a field of characteristic 0 is separable. A polynomial over such a field is separable if and only if it is the product of distinct irreducible polynomials.

Proof. Since every field of characteristic 0 contains $\mathbb{Q}$ as a subfield, it suffices to focus on $\mathbb{Q}$. Let $f(x) \in \mathbb{Q}[x]$ be irreducible and $\alpha \in \mathbb{C}$ be its root with multiplicity $k > 1$. Then
$$f(x) = (x-\alpha)^k g(x) \tag{601}$$
for some $g(x)$. Taking the derivative we have
$$f'(x) = k(x-\alpha)^{k-1} g(x) + (x-\alpha)^k g'(x) = (x-\alpha)^{k-1}\big(kg(x) + (x-\alpha)g'(x)\big) \tag{602}$$
which also has α as a root. So since $(x-\alpha)$ divides both f and f', it divides $\gcd(f, f')$.

By estabalishing a bit of theory of separability, can now detour here to find a classification of *all* finite fields, which is quite a powerful result. We have all the tools needed for this. We know that a field—as an integral domain—has characteristic 0 or prime p. We also know that a field is a vector space, at least over itself. But now that we have shown that a field can be modeled as a vector space, we can apply this to finite fields.

Theorem 11.19 (Characteristic Determines Base Field of Vector Space)

Given a field F,
1. If $\text{char}(F) = p$, then F is a vector space over $\mathbb{Z}_p$.
2. If $\text{char}(F) = 0$, then F is a vector space over $\mathbb{Q}$.

Proof.

Definition 11.10 (Frobenius Endomorphism)

Let F_p be a field of characteristic p. Then the map

$$\phi : F_p \to F_p, \qquad \phi(a) = a^p \tag{603}$$

is an injective field endormophism, called the **Frobenius endomorphism**. That is, it satisfies the following.

$$(a+b)^p = a^p + b^p, \qquad (ab)^p = a^p b^p \tag{604}$$

which is often called the *Freshman's dream*.

Proof. We prove the following properties.
1. *Addition.* We have
$$(a+b)^p = \sum_{k=0}^{p} \binom{p}{k} a^{p-k} b^k \tag{605}$$
It is clear that
$$\binom{p}{k} = \frac{p(p-1)...(p-k+1)}{k!} \tag{606}$$
is divisible by p for all $k \neq 0, p$, so all the middle terms must cancel out to 0.
2. *Multiplication.*
3. *Injectivity.*

Therefore, just from the characteristic we can classify all fields as vector spaces over either $\mathbb{Q}$ or $\mathbb{Z}_p$. Now if we focus on finite fields, we can do a reverse classification.

Theorem 11.20 (Finite Fields Have Cardinality p^d)

Let F be a finite field. Then $|F| = p^n$ for some $n \in \mathbb{N}$.

Proof. F is a vector space over $\mathbb{Z}_p$ from 11.19. Since F has finitely many elements, F has a finite spanning set, which implies $\dim_{\mathbb{Z}_p} F \leq +\infty$. Let d be the dimension and $\{b_1, \dots, b_d\}$ be the basis. The elements of F are

$$a_1 b_1 + \ldots + a_d b_d \tag{607}$$

with $a_1, \dots a_d \in \mathbb{Z}_p$. Thus there are p^d elements of F, so $F \simeq \mathbb{Z}_p^d$.

In fact, for *every* prime power there exists a unique field. Therefore we can create a bijection by proving the converse.

Theorem 11.21 (Field for Every p^d)

For every prime p and $n \in \mathbb{N}$, there exists a field with $q = p^d$ elements, unique up to isomorphism.

Proof. Let $f(x) = x^q - x \in \mathbb{Z}_p[x]$. Then this polynomial has a splitting field $K \supset \mathbb{Z}$. Now we claim the roots of $f(x)$ in K are distinct and form a subfield $F_q \subset K$. This will complete the proof since $F_q \subset K$ and $K \subset F_q \implies K = F_q$. Assume $\alpha, \beta \in K$ are roots of $f(x)$, and so $\alpha^p = \alpha$ and $\beta^p = \beta$

1. $\alpha + \beta \in K$ since by a modification of Freshman's dream, $(\alpha + \beta)^p = \alpha^p + \beta^p = \alpha + \beta$.[a]
2. $(-\alpha)^q = (-1)^q \alpha^q = (-1)^q \alpha = -\alpha$ since $-1 = 1$ or q is odd.
3. $\alpha\beta \in K$ since $\mathbb{Z}_p$ is a field and so $(\alpha\beta)^p = \alpha^p \beta^p = \alpha\beta$.
4. For multiplicative inverses, let $\alpha \neq 0$. Then $(\alpha^{-1})^p = (\alpha^p)^{-1} = \alpha^{-1}$.
5. For all p, 0 and 1 are roots so $0, 1 \in K$.

Now we show that K consists of distinct roots. Certainly $0 \in K$ with multiplicity 1 since $f(x) = x(x^{q-1} - x)$. Now suppose nonzero $r \in K$ is a root with multiplicity m. The multiplicity of r is the multiplicity of 0 of

$$f(x + r) = (x + r)^q - (x + r) = x^q + r^q - x - r = x^q - x \tag{608}$$

where the final step follows from $0 = r^q - r$ since $r \in K$. Therefore r has multiplicity 1. Since $K[x]$ has unique factorization property, it follows that $m = 1$ and every r is a simple root.

To show that every field with p^n elements is unique, let F be such a field. We claim that $\text{char}(F) = p \implies \mathbb{Z}_p \subset F$. We claim that every element of F is a root of $f(x) = x^q - x \in \mathbb{Z}_p[x]$, where F is the splitting field. Let $G = F^*$ be the multiplicative group of units. Since F is a field, then $|F^*| = |F| - 1 = p^d - 1$, and by constructing the cyclic group $\langle g \rangle \subset G$ for any $g \in G$, we know by Lagrange's theorem that $g^{|G|} = 1_G$, which implies that for all $x \in F$,

1. If $x \neq 0$ then $x^{p^d - 1} = x \implies x^{p^d} = x$ and so $x \in K$.
2. If $x = 0$ then $x^{p^d} - x = 0$ and so $x \in K$.

Therefore $F \subset K$ with $|F| = |K|$ both finite, and so $F = K$.

[a] We induct on n for $q = p^n$. For $n = 1$, this is trivial by Freshmans dream. Now assume it holds for some $n \in \mathbb{N}$. Then $(x + y)^{p^{n+1}} = ((x + y)^{p^n})^p = (x^{p^n} + y^{p^n})^p = (x^{p^n})^p + (y^{p^n})^p = x^{p^{n+1}} + y^{p^{n+1}}$.

From this, we can write for every prime p and natural n the finite field of order p^n as $\mathbb{F}_{p^n}$. It is clear that if $n = 1$ then $\mathbb{F}_p \simeq \mathbb{Z}_p$. The final result we will show is a hierarchy of subfields.

Theorem 11.22 (Hierarchy of Fields)

For a given prime p, if $p^m < p^n$, then

$$F_{q^m} \subset F_{q^n} \iff m \mid n \tag{609}$$

11.5 Automorphism Groups

Now we delve into the heart of Galois theory, which considers the relation of the group of permutations of the roots of $f(x)$ to the algebraic structure of its splitting field. The connection is given by the fundamental theorem of Galois theory.

As we stated in the beginning of this section, we want to look at symmetries of the roots of a polynomial. More concretely, given a polynomial $f(x) \in F[x]$, we can construct its splitting field K. TBD

More concretely, rather than directly computing all roots $\alpha_1, \dots, \alpha_n$ of a polynomial $f(x) \in F[x]$—which may be extremely hard—we can try to look at a certain transformation group of its permutations—that is, it's symmetries. But it's not just simply

Once we know that field extensions are vector spaces, what constitutes linear maps? A first guess would be

a ring homomorphism, but this may not be true. Therefore, we need some additional constraint.

Definition 11.11 (Field Automorphisms)

Let K/F be a field extension.

1. An **F-automorphism of K** is a ring homomorphism $\sigma : K \to K$ such that $\sigma(a) = a$ for all $a \in F$.[a]
2. The set of all F-automorphisms of K under composition is a subgroup of the autormorphism group of K, called the **F-automorphism group of K** and denoted $\text{Aut}(K/F)$.[b]

[a] Note that we denote it σ since it is analogous to a permutation—as we will see soon.

[b] Sometimes—in bad taste—this is introduced as the Galois group, but technically we need some extra conditions. To minimize confusion, I will refer to this as the automorphism group.

Proof. This is indeed a group under composition. The identity map $\iota \in \text{Aut}(K/F)$. The composition is clearly closed. Now given that $\sigma \in \text{Aut}(K/F)$, ϕ^{-1} is also an automorphism that is constant on F, so it is also in $\text{Aut}(K/F)$.

Lemma 11.23 (Linear Map)

An F-automorphism of K $\sigma : K \to K$ is a linear map of F-vector spaces.

Proof.

Essentially, the F-automorphism group is a transformation subgroup of the ring automorphism group of K that doesn't vary $F \subset K$. Let's provide a few examples to derive some of the automorphism groups.

Example 11.20 (Computing $\text{Aut}(\mathbb{Q}(\sqrt{2})/\mathbb{Q})$)

Given the field extension $\mathbb{Q} \subset \mathbb{Q}(\sqrt{2})$, we have for any $a, b \in \mathbb{Q}$ and $\phi \in \text{Aut}(\mathbb{Q}(\sqrt{2})/\mathbb{Q})$, we have

$$\phi(a + b\sqrt{2}) = \phi(a) + \phi(b\sqrt{2}) = a + b\phi(\sqrt{2}) \tag{610}$$

So ϕ is completely determined by the value of $\phi(\sqrt{2})$. Now let $\phi(\sqrt{2}) = \alpha + \beta\sqrt{2}$ for some $\alpha, \beta \in \mathbb{Q}$. We have

$$2 = \phi(2) = \phi(\sqrt{2}\sqrt{2}) = \phi(\sqrt{2})^2 = (\alpha + \beta\sqrt{2})^2 = (\alpha^2 + 2\beta^2) + 2\alpha\beta\sqrt{2} \tag{611}$$

Therefore $\alpha\beta = 0$ and $\alpha^2 + 2\beta^2 = 2$. With further casework, we must have $\alpha = 0, \beta = \pm 1$. In conclusion, there are exactly two $\mathbb{Q}$-automorphisms of $\mathbb{Q}(\sqrt{2})$.

1. The identity map $\iota(a + b\sqrt{2}) = a + b\sqrt{2}$, and
2. The conjugation map $\phi(a + b\sqrt{2}) = a - b\sqrt{2}$.

Example 11.21 (Computing $\text{Aut}(\mathbb{Q}(2^{1/3})/\mathbb{Q})$)

Given the field extension $\mathbb{Q} \subset \mathbb{Q}(\sqrt[3]{2})$, let us write $\xi = \sqrt[3]{2}$. Then since $\mathbb{Q}(\xi)$ is a vector space with basis $1, \xi, \xi^2$, we can write any element as $a + b\xi + c\xi^2$, and so by definition an element $\phi \in \text{Aut}(\mathbb{Q}(\xi)/\mathbb{Q})$ must satisfy

$$\phi(a + b\xi + c\xi^2) = a + b\phi(\xi) + c\phi(\xi)^2 \tag{612}$$

So the action is completely determined by the value of $\phi(\xi)$. Suppose $\phi(\xi) = \alpha + \beta\xi + \gamma\xi^2$ for some

$\alpha, \beta, \gamma \in \mathbb{Q}$. Through some derivation we have

$$2 = \phi(2) = \phi(\xi^3) = (\phi(\xi))^3 = (\alpha + \beta\xi + \gamma\xi^2)^3 \tag{613}$$
$$= (\alpha^3 + 2\beta^3 + 4\gamma^3) + 3(\alpha^2\beta + \beta^2\gamma + \gamma^2\alpha)\xi + 3(\alpha\beta^2 + \beta\gamma^2 + \gamma\alpha^2)\xi^2 \tag{614}$$

From the linear independence of $1, \xi, \xi^2$ we can see that

$$\alpha^3 + 2\beta^3 + 4\gamma^3 = 2 \tag{615}$$
$$\alpha^2\beta + \beta^2\gamma + \gamma^2\alpha = 0 \tag{616}$$
$$\alpha\beta^2 + \beta\gamma^2 + \gamma\alpha^2 = 0 \tag{617}$$

which turns out to have the only solution $\alpha = \gamma = 0, \beta = 1$. Therefore, the only $\mathbb{Q}$-automorphism of $\mathbb{Q}(\sqrt[3]{2})$ is the idnetity map.

Now let's look at how an F-automorphism acts on the roots of a polynomial. Let $f(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_1x + a_0 \in F[x]$, $F \subset K$ a field extension, and suppose $f(x)$ has roots $\alpha_1, \ldots, \alpha_m$ lying in K (and perhaps other roots lying in a further extension). It turns out that any F-automorphism of K must permute $\alpha_1, \ldots, \alpha_m$, so the roots stay "within" the polynomial.

Lemma 11.24 (F-Automorphims Permute Roots)

Let K/F be a field extension and let $\sigma \in \text{Aut}(K/F)$.

1. If $\alpha \in K$ is algebraic over F—i.e. there exists $f(x) \in F[x]$ s.t. $f(\alpha) = 0$—then $\sigma(\alpha) \in K$ is also a root of $f(x)$.[a]
2. Let $f(x) \in F[x]$ have roots $\alpha_1, \ldots, \alpha_m \in K$ (with possibly other roots outside of K). Then there exists a well-defined group homomorphism
$$\phi : \text{Aut}(K/F) \to S_m, \qquad \phi(\sigma)(i) \coloneqq j \text{ s.t. } \sigma(\alpha_j) = \alpha_i \tag{618}$$
3. If K is the splitting field of $f(x)$, then ϕ is injective and so $\text{Im}(\phi)$ is a subgroup of S_m.

[a] However it may not be a permutation! The next point describes this a bit more precisely.

Proof. Listed.

1. We know that $f(\alpha) = 0$. Now we can use the ring homomorphism properties to find
$$\sigma\big(f(\alpha)\big) = \sigma\big(a_n\alpha^n + a_{n-1}\alpha^{n-1} + \ldots + a_1\alpha + a_0\big) \tag{619}$$
$$= \sigma(a_n)\sigma(\alpha)^n + \sigma(a_{n-1})\sigma(\alpha)^{n-1} + \ldots + \sigma(a_1)\sigma(\alpha) + \sigma(a_0) \tag{620}$$
$$= a_n\sigma(\alpha)^n + a_{n-1}\sigma(\alpha)^{n-1} + \ldots + a_1\sigma(\alpha) + a_0 \tag{621}$$
which implies that $f(\sigma(\alpha)) = 0$.
2. Now assume that $S = \{\alpha_1, \ldots, \alpha_n\}$ is contained in K. We know that $\phi(\alpha_j) \in S$. The map $\phi \mapsto \phi(\alpha_j)$ is indeed a group homomorphism $\text{Aut}(K/F) \to \text{Perm}(S)$.
3. If K is the splitting field, then $K = F[S]$, and so if $\phi \in \text{Aut}(K/F)$ fixes all the α_j's, then it must fix all of K.

Example 11.22 (Recomputing $\text{Aut}(\mathbb{Q}(2^{1/3})/\mathbb{Q})$)

Therefore, we can get a much simpler solution of the Automorphism group of $\mathbb{Q}(\sqrt[3]{2})$ over $\mathbb{Q}$. Since $\xi = \sqrt[3]{2}$ is a root of the irreducible polynomial $x^3 - 2$, any $\phi \in \text{Aut}(\mathbb{Q}[\xi]/\mathbb{Q})$ must carry ξ to some other root of $x^3 - 2$. The other roots are in the complex plane and not in $\mathbb{Q}(\sqrt[3]{2})$, and so ϕ must carry $\xi \mapsto \xi$.

Thus, ϕ must be the identity.

Example 11.23 (Computing $\mathrm{Aut}(\mathbb{Q}(1+\sqrt{2})/\mathbb{Q})$)

Let $f(x) = x^2 - 2x - 1 \in \mathbb{Q}[x]$, which is the minimal polynomial of $1+\sqrt{2} \in \mathbb{R}$. One root of $f(x)$ is $1+\sqrt{2} \in \mathbb{Q}(\sqrt{2})$. But we know that there exists a $\phi \in \mathrm{Aut}(\mathbb{Q}(\sqrt{2})/\mathbb{Q})$ that conjugates, and so $1-\sqrt{2}$ must also be a root.

Example 11.24 (Computing $\mathrm{Aut}(\mathbb{C}/\mathbb{R})$)

It follows that $\mathrm{Aut}(\mathbb{C}/\mathbb{R})$ is a group of order 2 generated by complex conjugation.

So given any field extension K/F, we can associate it with a group $\mathrm{Aut}(K/F)$. One can also reverse this process and associate to each group of automorphisms a field extension. In fact, the subgroup property is not ecessary, and we can just talk about subsets.

Theorem 11.25

Let K be a field, and let H be a subset of the group of automorphisms $\mathrm{Aut}(K)$. Then the collection F of element of K fixed by all the elements of H is a subfield of K.

Proof. Let $h \in H$ and let $a, b \in F$. Then by definition $h(a) = a, h(b) = b$, and so $h(a+b) = h(a)+h(b)$, $h(ab) = h(a)h(b)$, and $h(a)^{-1} = a^{-1}$. So F is closed, hence a subfield of K.

Theorem 11.26 (Inclusion Reversing Association between Groups and Fields)

Therefore, the association of groups to fields and fields to group are inclusion reversing.

1. If $F_1 \subset F_2 \subset K$ are two subfields of K, then $\mathrm{Aut}(K/F_2) \subset \mathrm{Aut}(K/F_1)$.
2. If $H_1 \subset H_2 \subset \mathrm{Aut}(K)$ are to subgroups of automorphisms of K with fixed fields F_1 and F_2, respectively, then $F_2 \subset F_1$.

Proof.

Given a subfield F of K, the associated group is the collection of F-automorphisms of K. Given a group of F-automorphisms of K, the associated extension is defined by taqking F to be the fixed field of the automorphisms.

1. Given the subfield $\mathbb{Q} \subset \mathbb{Q}(\sqrt{2})$, the automorphism group is $\{1, \sigma\}$, and given the group, the fixed field is $\mathbb{Q}$.
$$\mathbb{Q}(\sqrt{2}) \supset \mathbb{Q} \longrightarrow \{1, \sigma\} \longrightarrow \mathbb{Q} \tag{622}$$
Therefore, there is a duality between the subfield $\mathbb{Q}$ and the group $\{1, \sigma\}$.
2. Given the subfield $\mathbb{Q} \subset \mathbb{Q}(\sqrt[3]{2})$, we obtain the trivial group $\{e\}$, which induces the fixed field $\mathbb{Q}(\sqrt[3]{2})$.
$$\mathbb{Q}(\sqrt[3]{2}) \supset \mathbb{Q} \longrightarrow \{e\} \longrightarrow \mathbb{Q}(\sqrt[3]{2}) \tag{623}$$
In here, we lose the duality.

Let's investigate why we lose the duality for the second example. The trivial group $\{e\}$ does not have enough automorphisms to force the fixed field to be $\mathbb{Q}$ rather than the full $\mathbb{Q}(\sqrt[3]{2})$. This is because the other roots of the minimal polynomial $x^3 - 2 \in \mathbb{Q}[x]$—which can be the images of $\sqrt[3]{2}$ under an automorphism—lie outside

of $\mathbb{Q}(\sqrt[3]{2})$. We now make precise the notion of fields with "enough" automorphisms. Essentially we want to count the number of elements in the automorphism group. How do we do this? Well we take a field extension K/F, take the identity automorphism of F, and see how many ways we can extend it into an automorphism of K.

Recall 11.11, which states that any isomorphism $\phi : F \to F'$ can be extended to an isomorphism $\bar{\phi} : K \to K'$ for $f'(x) = \phi(f(x)) \in F'[x]$. We build on this theorem.

Theorem 11.27 (Bound on Number of Field Isomorphism Extensions)

Let $F \subset K, F' \subset K'$ be field extensions with an isomorphism $\phi : F \to F'$. Then, the numbers of extensions $\bar{\phi} : K \to K'$ is bounded by $[K : F]$.

Proof. We now show by induction on $[K : F]$ that the number of such extensions is at most $[K : F]$, with equality if $f(x)$ is separable over F. If $[K : F] = 1$ then $K = F$, $K' = F'$, $\sigma = \varphi$ and the number of extensions is 1. If $[K : F] > 1$ then $f(x)$ has at least one irreducible factor $p(x)$ of degree > 1 with corresponding irreducible factor $p'(x)$ of $f'(x)$. Let α be a fixed root of $p(x)$. If σ is any extension of φ to K, then σ restricted to the subfield $F(\alpha)$ of K is an isomorphism τ of $F(\alpha)$ with some subfield of K'. The isomorphism τ is completely determined by its action on α, i.e., by $\tau\alpha$, since α generates $F(\alpha)$ over F. Just as in Proposition 2, we see that $\tau\alpha$ must be some root β of $p'(x)$. Then we have a diagram

$$\begin{array}{ccc} K & \xrightarrow{\sigma} & K' \\ \downarrow & & \downarrow \\ F(\alpha) & \xrightarrow{\tau} & F'(\beta) \\ \downarrow & & \downarrow \\ F & \xrightarrow{\varphi} & F' \end{array}$$

Conversely, for any β a root of $p'(x)$ there are extensions τ and σ giving such a diagram (this is Theorem 13.8 and Theorem 13.27). Hence to count the number of extensions σ we need only count the possible number of these diagrams.
The number of extensions of φ to an isomorphism τ is equal to the number of distinct roots β of $p'(x)$. Since the degree of $p(x)$ and $p'(x)$ are both equal to $[F(\alpha) : F]$, we see that the number of extensions of φ to a τ is at most $[F(\alpha) : F]$, with equality if the roots of $p(x)$ are distinct.
Since K is also the splitting field of $f(x)$ over $F(\alpha)$, K' is the splitting field of $f'(x)$ over $F'(\beta)$, and $[K : F(\alpha)] < [K : F]$, we may apply our induction hypothesis to these field extensions. By induction, the number of extensions of τ to σ is $\leq [K : F(\alpha)]$, with equality if $f(x)$ has distinct roots.
From $[K : F] = [K : F(\alpha)][F(\alpha) : F]$ it follows that the number of extensions of φ to σ is $\leq [K : F]$. We have equality if $p(x)$ and $f(x)$ have distinct roots, which is equivalent to $f(x)$ having distinct roots since $p(x)$ is a factor of $f(x)$, completing the proof by induction.

Corollary 11.28 (Bound on Order of Automorphism Group)

Let K be the splitting field of a polynomial $f(x) \in F[x]$. Then,

$$|\operatorname{Aut}(K/F)| \leq [E : F] \tag{624}$$

with equality if $f(x)$ is separable over F.

11.6 Galois Extensions and Galois Groups

If we reach this bound, then this is equivalent to saying that the automorphism group H has "enough" elements to recover F (given field extension K/F) as the fixed field of H.

Definition 11.12 (Galois Field Extension)

A field extension $F \subset K$ is **Galois** if $|\operatorname{Aut}(K/F)| = [K : F]$—or equivalently, the degree of the minimal polynomial $f(x) \in F[x]$ that splits in K.

Definition 11.13 (Galois Group of Galois Field Extension)

Given a Galois field extension K/F, the automorphism group $\operatorname{Aut}(K/F)$ is called the **Galois Group**, denoted $\mathcal{G}(K/F)$.

Therefore, we can think of an automorphism group as Galois if we have reached the maximal number of automorphisms.

Example 11.25 (Simple Examples)

We review the derived Galois groups above and see if the field extensions are Galois.

1. $\mathbb{Q}(\sqrt{2})$ is a Galois extension of $\mathbb{Q}$. Since we know that $G = \mathcal{G}(\mathbb{Q}(\sqrt{2})/\mathbb{Q})$ has order 2, which is the same as the dimension of $\mathbb{Q}(\sqrt{2})$, i.e. the degree of the minimal polynomial $x^2 - 2 \in \mathbb{Q}[x]$.
2. $\mathbb{Q}(\sqrt[3]{2})$ is not a Galois extension of $\mathbb{Q}$ since $|\mathcal{G}(\mathbb{Q}[\sqrt[3]{2}]/\mathbb{Q})| = 1$ but $[\mathbb{Q}[\sqrt[3]{2}] : \mathbb{Q}] = 3$.
3. Let $K = \mathbb{Q}[\sqrt[3]{2}, i\sqrt{3}]$ is the splitting field of $f(x) = x^3 - 2$. Then $[K : \mathbb{Q}] = 6$. $\mathcal{G}(K/\mathbb{Q}) \simeq S_3$ has order 6 since we showed 6 $\mathbb{Q}$-automorphisms of K, but now we know that there can be no more.
4. Let $\alpha = \sqrt[7]{2}$. $\mathbb{Q} \subset \mathbb{Q}[\alpha]$ is not a Galois extension. $x^7 - 2$ is a polynomial with root α, and by Eisenstein $x^7 - 2$ is irreducible. So $f(x) = x^7 - 2$ is the minimal polynomial of α. This means that the number of $\mathbb{Q}$-embeddings $\mathbb{Q}[\alpha] \to \mathbb{Q}[\alpha]$ is in bijection with the number of roots of $f(x)$ in $\mathbb{Q}[\alpha]$. All 7 roots of $f(x)$ are $\sqrt[7]{2}e^{2\pi ij/7}$ for $j = 0, \dots, 6$, which has one real root. So there is 1 $\mathbb{Q}$-embedding $\mathbb{Q}[\alpha] \to \mathbb{Q}[\alpha]$. But $[\mathbb{Q}[\sqrt[7]{2}] : \mathbb{Q}] = \deg f(x) = 7$. Since $\mathbb{Q}[\alpha] \simeq \mathbb{Q}[x]/\langle x^7 - 2 \rangle$, which are polynomials of degree at most 6.

These examples suggest that K will be a Galois extension of F whenever K is a splitting field of some polynomial $f(x) \in F[x]$. To establish this, we must produce $[K : F]$ F-automorphisms of K under these circumstances.

Theorem 11.29 (Splitting Fields of Separable Polynomials are Galois)

If K is the splitting field over F of a separable polynomial $f(x)$ then K/F is a Galois extension.

Proof. Since K is the splitting field of $f(x)$ over F, we have $F[r_1, \dots, r_m]$ where the r_i are the distinct roots of $f(x)$. We show by induction that there are $[F[r_1, \dots, r_j] : F]$ F-embeddings of $F[r_1, \dots, r_j] \to K$. For $j = 1$, r_1 is a root of some irreducible factor $f_1(x)$ of $f(x)$.

$$F[r_1] \simeq \frac{F[x]}{\langle f_1(x) \rangle} \tag{625}$$

and the set of F-embeddings $F[r_1] \to K$ is in bijection with the set of roots $\alpha \in K$ of $f_1(x)$. By hypothesis , $f(x)$ has no repeated roots, which implies that the number of F-embeddings $F[r_1] \to K$ is $\deg f_1(x) = [F[r_1] : F]$ which gives the base case. For the inductive step, we know there are

$$[F[r_1, \dots, r_{j-1}] : F] \tag{626}$$

F-embeddings $F[r_1, \dots, r_{j-1}] \xrightarrow{\phi} K$. For each, we will show that there are exactly $[F[r_1, \dots, r_j] : F[r_1, \dots, r_{j-1}]]$ extensions of ϕ which completes the proof. Because

$$F[r_1, \dots, r_j] : F] = F[r_1, \dots, r_j] : F[r_1, \dots, r_{j-1}]]F[r_1, \dots, r_{j-1}] : F] \tag{627}$$

Let $g(x)$ be the minimal polynomial of r_j over $F[r_1, \dots, r_{j-1}]$. Since $g(r_j) = 0$, g divides one of the irreducible factors of $f(x)$ in $F[r_1, \dots, r_{j-1}][x]$ which implies it has no repeated roots. Then

$$F[r_1, \dots, r_j] = \frac{E[x]}{\langle g(x) \rangle} \tag{628}$$

The number of E-embeddings is equal to the number of roots in in g which is $\deg g = F[r_1, \dots, r_j : F]$.

It turns out that the converse of the theorem is also true, as we will see later. This gives us a complete characterization of Galois extensions. The following is immediate.

Corollary 11.30 (Splitting Fields of $\mathbb{Q}$ are Automatically Galois)

If K is a splitting field of $f(x) \in \mathbb{Q}[x]$, then $K/\mathbb{Q}$ is Galois.

Proof. All irreducible factors of $f(x) \in \mathbb{Q}[x]$ has no repeated roots and hence is separable from 11.18. Therefore, from 11.29, since K is the splitting field, $K/\mathbb{Q}$ is Galois.

What about the converse? There are two steps to proving that for any Galois field extension $F \subset K$, there exists a polynomial $f(x) \in F[x]$ that splits in K.

Lemma 11.31 (Irreducible Polynomial with Root in Galois Extension Splits)

Let $F \subset K$ be a Galois extension of fields. Let $f(x) \in F[x]$ be an irreducible polynomial with a root $\alpha \in K$. Then $f(x)$ splits in K, i.e. all other roots must be in K.

Proof. Let σ_j for $j = 1, \dots, n$ be the elements of $\mathcal{G}(K/F)$, and set $\alpha_j = \sigma_j(\alpha)$. Then define

$$h(x) = \prod_{i=1}^{n} (x - \alpha_i) \in K[x] \tag{629}$$

We claim that the coefficients of $h(x)$ are fixed by any element of $\mathcal{G}(K/F)$. This is because by 11.24, the coefficients obtained by taking the binomial expansion stay invariant. Therefore, all coefficients of $h(x)$ are in F, and so $h(x) \in F[x]$. Since $g(x)$ and $h(x)$ have a common root $\alpha \in K$, and since $g(x)$ is irreducible in $F[x]$, we have $g(x) \mid h(x)$. Therefore $g(x)$ as a factor of $h(x)$ which splits in K, also splits in $K[x]$.

Theorem 11.32 (Every Galois Extension has a Splitting Field)

Let $F \subset K$ be a Galois extension. Then K is the splitting field of a polynomial $f(x) \in F[x]$.

Proof. Let $\alpha_1, \dots, \alpha_k$ be a basis for K over F, and for each $j = 1, \dots, k$, let $g_j(x) \in F[x]$ be an irreducible polynomial with root α_j. Then by 11.31, K is the splitting field $g_j(x)$ splits in K, which

implies that

$$f(x) = g_1(x) \dots g_k(x) \in F[x] \tag{630}$$

also splits in K, and it does in no smaller fields since any splitting field must contain all the α_j's.

When computing Galois groups, it only makes sense to talk about it with respec to Galois field extensions. Therefore, if we are given a field extension, we must prove that it is a Galois extension, then find the order of its Galois group, then compute the elements using the theorems we have above. But since we have established equivalent conditions for a Galois group to exist through polynomials, we can talk about a Galois group w.r.t. a separable polynomial.

Definition 11.14 (Galois Group of Separable Polynomial)

Given a separable polynomial $f(x) \in F[x]$, its **Galois group** is defined to be the Galois group of the splitting field of $f(x)$ over F.

If we are given a polynomial $f(x) \in F[x]$, when we must find that no irreducible factor of it has repeated roots, and then we find the splitting field F/K to compute the Galois group. Let's do some practice.

Example 11.26 (Computing Galois Groups)

Now let's compute the Galois group of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ over $\mathbb{Q}$.

1. $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ is the splitting field of $f(x) = (x^2 - 2)(x^2 - 3)$ over $\mathbb{Q}$. Then the field extension $\mathbb{Q} \subset \mathbb{Q}(\sqrt{2}, \sqrt{3})$ is Galois.
2. Therefore the Galois group of this extension exists. Since $\sqrt{2}, \sqrt{3}$ are not linearly dependent, we have $[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}] = 4$. Thus the Galois group has order 4. Now it remains to compute the 4 elements.
3. Since $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ is a $\mathbb{Q}$-vector space of dimension 4, we know that every element is of the form $a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}$. Since

$$\phi(a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}) = a + b\phi(\sqrt{2}) + c\phi(\sqrt{3}) + d\phi(\sqrt{2})\phi(\sqrt{3}), \tag{631}$$

   every $\mathbb{Q}$-automorphism is determined by the values $\phi(2), \phi(3)$.
4. From 11.24, we know that elements $\phi \in \mathcal{G}(\mathbb{Q}(\sqrt{2}, \sqrt{3})/\mathbb{Q})$ permute the roots of $f(x)$. We immediately see that if $\phi(\sqrt{2}) = \sqrt{3}$

$$2 = \phi(2) = \phi(\sqrt{2})^2 = \sqrt{3}^2 = 3 \tag{632}$$

   which is a contradiction. The same logic follows for $\sqrt{2} \mapsto \pm\sqrt{3}, \sqrt{3} \mapsto \pm\sqrt{2}$. So it must be the case that $\phi(\sqrt{2}) = \pm\sqrt{2}$ and $\phi(\sqrt{3}) = \pm\sqrt{3}$.

Therefore, we are able to deduce that the 4 automorphisms are.

$$\phi_1(a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}) = a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6} \tag{633}$$

$$\phi_1(a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}) = a - b\sqrt{2} + c\sqrt{3} - d\sqrt{6} \tag{634}$$

$$\phi_1(a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}) = a + b\sqrt{2} - c\sqrt{3} - d\sqrt{6} \tag{635}$$

$$\phi_1(a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}) = a - b\sqrt{2} - c\sqrt{3} + d\sqrt{6} \tag{636}$$

which is isormophic to the Klein 4-group.

Example 11.27 (Computing Galois Groups of 3 Adjoining Elements)

For $\mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5})$, follow the same method, as it's the splitting field of $(x^2-2)(x^2-3)(x^2-5)$. Consider the automorphisms:

$$\sqrt{2} \mapsto -\sqrt{2} \tag{637}$$
$$\sqrt{3} \mapsto \sqrt{3} \tag{638}$$
$$\sqrt{5} \mapsto \sqrt{5} \tag{639}$$

and

$$\sqrt{2} \mapsto \sqrt{2} \tag{640}$$
$$\sqrt{3} \mapsto -\sqrt{3} \tag{641}$$
$$\sqrt{5} \mapsto \sqrt{5} \tag{642}$$

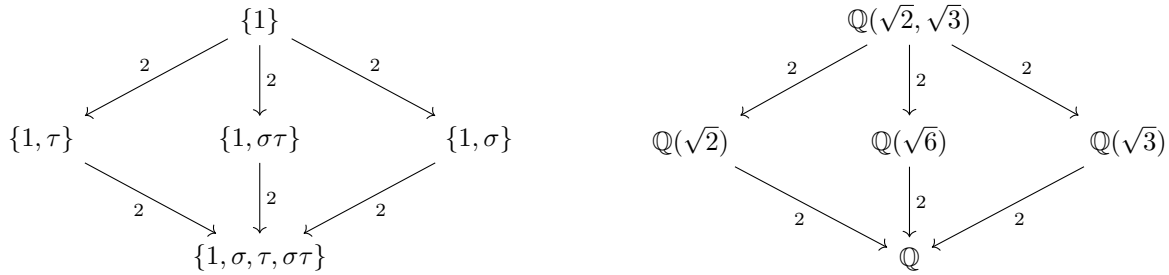
and

$$\sqrt{2} \mapsto \sqrt{2} \tag{643}$$
$$\sqrt{3} \mapsto \sqrt{3} \tag{644}$$
$$\sqrt{5} \mapsto -\sqrt{5} \tag{645}$$

which generates the group. We can see that this is $\mathbb{Z}_2^3$.

11.7 Fundamental Theorem of Galois Theory

Notice from our previous example that there seems to be a similarity between the known subgroups of the Klein-4 groups and the known subfields of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$. In fact this bijection is not a coincidence, and this is precisely the statement of the fundamental theorem of Galois theory.

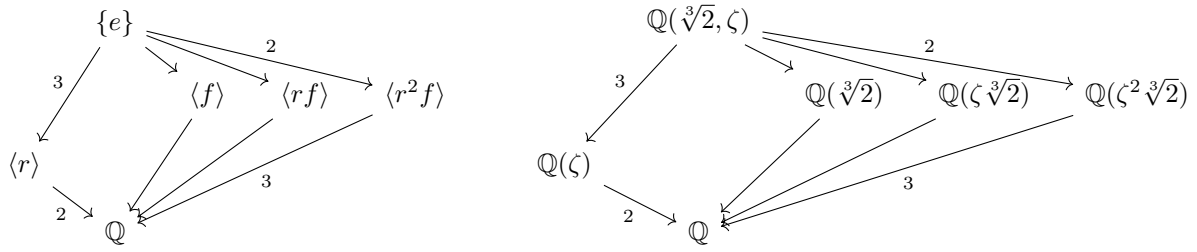


(a) Group structure diagram

(b) Field extension diagram

Figure 19: Lattice diagrams showing group structure and corresponding field extensions

Also take a look at this for the polynomial $x^3 - 2 \in \mathbb{Q}[x]$.



(a) (b)

Figure 20

Theorem 11.33 (Fundamental Theorem of Galois Theory)

Let K/F be a Galois extension and let $G = \mathcal{G}(K/F)$. Then there is a bijection

$$\{\text{Fields } E \mid F \subset E \subset K\} \simeq \{\text{Subgroups } H \mid \{e\} \subset H \subset G\} \tag{646}$$

where given automorphism group H, E is the fixed field of all elements in H; and given field E, H is the corresponding group formed by the elements of G fixing E. Under this correspondence, we have

1. If $E_1 \sim H_1, E_2 \sim H_2$, then $E_1 \subset E_2$ iff $H_2 \subset H_1$. As a consequence $F \simeq G$ and $K \simeq \{e\}$.
2. $[K : E] = |H|$ and $[E : F] = |G : H|$.[a]
3. K/E is always Galois, with Galois group

$$\mathcal{G}(K/E) = H \tag{647}$$

4. E is Galois over F iff H is a normal subgroup of G. If this is the case, then the Galois group is isomorphic to the quotient group

$$\mathcal{G}(E/F) \cong G/H \tag{648}$$

5. If E_1, E_2 correspond to H_1, H_2 respectively, then $E_1 \cap E_2 \sim \langle H_1, H_2 \rangle$ and the composite field $E_1E_2 \sim H_1 \cap H_2$. Hence the lattice of subfields of K containing F and the lattice of subgroups are "dual."

[a] Note the left is the degree of E over $\mathbb{Q}$ while the right is the number of cosets of H in G!

Proof.

Example 11.28

Since all the subgroups of an abelian group are normal, all the subfields of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ are Galois extensions of $\mathbb{Q}$.

11.8 Galois Groups of Polynomials

Definition 11.15 (Elementary Symmetric Functions)

Let $x_1, x_2, \dots, x_n$ be indeterminates. The **elementary symmetric functions** $s_1, s_2, \dots, s_n$ are defined

as

$$s_1 = x_1 + x_2 + \ldots + x_n \tag{649}$$
$$s_2 = x_1x_2 + x_1x_3 + \ldots + x_ix_j + x_{n-1}x_n \tag{650}$$
$$\vdots = \vdots \tag{651}$$
$$s_n = x_1x_2 \ldots x_n \tag{652}$$

Definition 11.16 (General Polynomial of Degree n)

The **general polynomial of degree** n is the polynomial

$$f(x) = \prod_{i=1}^{n}(x - x_i) \tag{653}$$

whose roots are the indeterminates $x_1, \ldots, x_n$.

Theorem 11.34

The fixed field of the symmetric group S_n acting on the field of rational functions in n variables $F(x_1, \ldots, x_n)$ is the field of rational functions in the eleemntary symmetric functions $F(s_1, \ldots, s_n)$.

Definition 11.17 (Symmetric)

A rational function $f(x_1, \ldots, x_n)$ is **symmetric** if it is not changed by any permutation of the variables $x_1, \ldots x_n$.

Theorem 11.35 (Fundamental Theorem on Symmetric Functions)

Any symmetric functions in the variables $x_1, x_2, \ldots, x_n$ is a rational function in the elementary symmetric functions $s_1, s_2, \ldots, s_n$.

Theorem 11.36

The general polynomial

$$x^n - s_1x^{n-1} + s_2x^{n-2} + \ldots + (-1)^n s_n \tag{654}$$

over the field $F(s_1, \ldots, s_n)$ is separable with Galois group S_n.

The well known discriminant of a quadratic equation

$$f(x) = ax^2 + bx + c \tag{655}$$

is known in the form $\nabla = b^2 - 4ac$. However, we will present it in a slightly different manner.

Definition 11.18

The **discriminant** $D(\varphi)$ of a quadratic polynomial

$$\varphi = a_0x^2 + a_1x + a_2 \in \mathbb{C}[x] \tag{656}$$

with $c_1, c_2 \in \mathbb{C}$ as its roots is defined

$$D(\varphi) = a_1^2 - 4a_0a_2 = a_0^2\left(\left(\frac{a_1}{a_0}\right)^2 - \frac{4a_2}{a_0}\right) = a_0^2\big((c_1+c_2)^2 - 4c_1c_2\big) = a_0^2(c_1-c_2)^2 \tag{657}$$

Clearly, the value of $D(\varphi)$ can tell us three things

1. $c_1, c_2 \in \mathbb{R}, c_1 \neq c_2$. Then $c_1 - c_2$ is a nonzero real number and $D(\varphi) > 0$.
2. $c_1 = c_2 \in \mathbb{R}$. Then $c_1 - c_2 = 0$ and $D(\varphi) = 0$.
3. $c_1, c_2 \in \mathbb{C}, c_1 = \bar{c}_2$. Then, $c_1 - c_2$ is a nonzero strictly imaginary number and $D(\varphi) < 0$.

Definition 11.19

We can generalize this notion of the discriminant to arbitrary polynomials

$$\varphi = a_0x^n + a_1x^{n-1} + ... + a_{n-1}x + a_n \in \mathbb{F}[x],\ a_0 \neq 0 \tag{658}$$

The discriminant $D(\varphi)$ of the polynomial above is defined

$$D(\varphi) \equiv a_0^{2n-2}\prod_{i>j}(c_i - c_j)^2 \tag{659}$$

The a_0 term isn't very important in this formula, since it does not affect whether $D(\varphi)$ is positive, negative, or zero.

Definition 11.20

A polynomial

$$\varphi = a_0x^n + a_1x^{n-1} + ... + a_{n-1}x + a_n \in \mathbb{F}[x],\ a_0 \neq 0 \tag{660}$$

where $a_1 = 0$ is called **depressed**. A depressed cubic polynomial is of form

$$\varphi = x^3 + px + q \tag{661}$$

Theorem 11.37

Every monic (leading coefficeint = 1) polynomial (and non-monic ones)

$$\varphi = x^n + a_1x^{n-1} + ... + a_{n-1}x + a_n \in \mathbb{F}[x],\ a_0 \neq 0 \tag{662}$$

can be turned into a depressed polynomial with the change of variable

$$x = y - \frac{a_1}{n} \tag{663}$$

to get the polynomial

$$\psi = y^n + b_2y^{n-2} + ... + b_{n-1}y + b_n \tag{664}$$

Lemma 11.38

A cubic polynomial

$$\varphi = a_0x^3 + a_1x^2 + a_2x + a_3 \in \mathbb{R}[x] \tag{665}$$

with roots $c_1, c_2, c_3 \in \mathbb{C}$ has discriminant

$$D(\varphi) \equiv a_0^4(c_1-c_2)^2(c_1-c_3)^2(c_2-c_3)^2 \tag{666}$$

With a bit of evaluation, it can also be expressed in terms of its coefficients as

$$D(\varphi) = a_1^2a_2^2 - 4a_1^3a_3 - 4a_0a_2^3 + 18a_0a_1a_2a_3 - 27a_0^2a_3^2 \tag{667}$$

Again, three possibilities can occur (up to reordering of its roots).

1. c_1, c_2, c_3 are distinct real numbers. Then $D(\varphi) > 0$.
2. $c_1, c_2, c_3 \in \mathbb{R}, c_1 = c_2$. Then $D(\varphi) = 0$.
3. $c_1 \in \mathbb{R}, c_2 = \bar{c}_3 \notin \mathbb{R}$. Then $D(\varphi) < 0$.

Furthermore, the cubic formula used to find the roots of the polynomial is

$$c_{1,2,3} = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{p^3}{27} + \frac{q^2}{4}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{p^3}{27} + \frac{q^2}{4}}} \tag{668}$$

known as **Cardano's formula**, after the mathematician Gerolamo Cardano.

11.9 Solvable and Radical Extensions

11.10 Exercises

Exercise 11.1 (Shifrin 5.3.3)

The polynomial $f(x) = x^2 + 1$ is irreducible in $\mathbb{Z}_3[x]$, and so $K = \mathbb{Z}_3[x]/(x^2 + 1)$ is a field with nine elements. Let $\alpha \in K$ be a root of $f(x)$. Find irreducible polynomials in $\mathbb{Z}_3[x]$ having as roots, respectively,

a. $\alpha + 1$
b. $\alpha - 1$.

Solution. Listed.

1. We can see that

$$\begin{aligned}(\alpha+1)^2 = \alpha^2 + 2\alpha + 1 = 2\alpha &\implies (\alpha+1)^2 - 2\alpha = 0 &(669)\\ &\implies (\alpha+1)^2 - 2\alpha - 2 + 2 = 0 &(670)\\ &\implies (\alpha+1)^2 - 2(\alpha+1) + 2 = 0 &(671)\end{aligned}$$

   and so $f(x) = x^2 - 2x + 2 \in \mathbb{Z}_3[x]$ has $\alpha + 1$ as a root.
2. Similarly, we have

$$\begin{aligned}(\alpha-1)^2 = \alpha^2 - 2\alpha + 1 = -2\alpha &\implies (\alpha-1)^2 + 2\alpha = 0 &(672)\\ &\implies (\alpha-1)^2 + 2\alpha - 2 + 2 = 0 &(673)\\ &\implies (\alpha-1)^2 + 2(\alpha-1) + 2 = 0 &(674)\end{aligned}$$

   and so $f(x) = x^2 + 2x + 2 \in \mathbb{Z}_3[x]$ has $\alpha - 1$ as a root.

Exercise 11.2 (Shifrin 5.3.4)

Construct explicitly an isomorphism

$$\mathbb{Z}_2[x]/(x^3 + x + 1) \to \mathbb{Z}_2[x]/(x^3 + x^2 + 1).$$

Solution. Both x^3+x+1 and x^3+x^2+1 are irreducible in $\mathbb{Z}_2[x]$, so both are fields of order 8 (since the x^3 is equivalent to a lower order polynomial) consisting of all polynomials in $\mathbb{Z}_2[x]$ of degree ≤ 2. We can construct the isomorphism ϕ sending $\phi(f(x)) = f(x+1)$. This is a homomorphism since it maps 1 to 1, and

$$\phi((f+g)(x)) = (f+g)(x+1) = f(x+1) + g(x+1) = \phi(f(x)) + \phi(g(x)) \tag{675}$$

$$\phi((fg)(x)) = (fg)(x+1) = f(x+1)g(x+1) = \phi(f(x))\phi(g(x)) \tag{676}$$

It is also bijective since the inverse mapping $\phi(f(x)) = f(x-1) = f(x+1)$ is well-defined. Finally, we can see that that considering ϕ as an automorphism over $\mathbb{Z}_2[x]$, $\phi(x^3+x+1) = (x+1)^3 + (x+1) + 1 = x^3 + x^2 + 1$, so it maps the ideals to each other. This therefore induces an isomorphism between the quotient rings. We can explicitly write out the image of each element.

1. $\phi(0) = 0$.
2. $\phi(1) = 1$.
3. $\phi(x) = x+1$
4. $\phi(x+1) = x$.
5. $\phi(x^2) = x^2+1$.
6. $\phi(x^2+1) = x^2$.
7. $\phi(x^2+x) = x^2+x$.
8. $\phi(x^2+x+1) = x^2+x+1$.

Exercise 11.3 (Shifrin 5.3.5)

Let F be a finite field of characteristic p. Show that every element $a \in F$ can be written in the form $a = b^p$ for some $b \in F$. (Hint: Consider the Frobenius automorphism.)

Solution. Then F has $q = p^n$ elements for some $n \in \mathbb{N}$, and in Shifrin we have established through Frobenius automorphism $\sigma(a) = a^p$ that $\sigma^n(a)$ is the identity, i.e.

$$a = \sigma^n(a) = (a^p)^n = a^{pn} = (a^n)^p \tag{677}$$

Therefore, we have found $b = a^n \in F$ satisfying the condition.

Exercise 11.4 (Shifrin 5.3.7)

Let $q = p^n$, and let $f(x) = x^q - x$.

a. Prove that if $g(x)$ is an irreducible polynomial of degree d in $\mathbb{Z}_p[x]$, then $g(x)$ divides $f(x)$ if and only if $d|n$.
b. Prove that $f(x)$ is the product of all monic, irreducible polynomials in $\mathbb{Z}_p[x]$ whose degrees divide n.

Solution. For (a), we prove bidirectionally. Since $g(x)$ is irreducible, $F = \mathbb{Z}_p[x]/\langle g(x) \rangle$ is a field of p^d elements and $g(x)$ is the minimal polynomial of α over $\mathbb{Z}_p$. We also know that for any element a in a field of order p^d, it satisfies $a^{p^d} = a$. Additionally, the multiplicative group of units $(\mathbb{Z}_p[x]/\langle g(x) \rangle)^*$ is a cyclic group of order $p^d - 1$ generated by α. By Lagrange's theorem, the order of any element of this multiplicative group must divide $p^d - 1$. Choosing α, we have $\alpha^{p^d-1} = 1 \implies \alpha^{p^d} = \alpha$.

1. ($\rightarrow$). Let $g(x)$ divide $f(x) = x^{p^n} - x$. Then $g(\alpha) = 0 \implies f(\alpha) = \alpha^{p^n} - \alpha = 0 \implies \alpha^{p^n} = \alpha$. Therefore,

$$\alpha = \alpha^{p^d} = \alpha^{p^n} \tag{678}$$

The smallest positive integer m such that $\alpha^{p^m} = \alpha$ is $m = d$ as $g(x)$ is the minimal polynomial.

Since $\alpha^{p^n} = \alpha$ and d is the smallest such exponent, we have $d \mid n$.

2. ($\leftarrow$). Assume that $d \mid n$. Consider the field $F = \mathbb{Z}_p[x]/\langle g(x)\rangle$, which is a field of order p^d. We also know that for any element a in a field of order p^d, it satisfies $a^{p^d} = a$. Taking $x \in \mathbb{Z}_p[x]$, its image $\bar{x} \in F$ has the property that $\bar{x}^{p^d} - \bar{x} = 0$, and so this means that $x^{q^d} - x$ is in the kernel of this quotient map. Therefore $(x^{p^d} - x) \in \langle g(x)\rangle \implies g(x) \mid (x^{p^d} - x)$. To prove the final step, we prove that $\forall d, n$, $x^{p^d} - x \mid x^{p^n} - x$ iff $d \mid n$.
Then we have $n = kd$ for some $k \in \mathbb{N}$, and so

$$\alpha^{p^n} = \alpha^{p^{kd}} = \alpha \tag{679}$$

and so α is a root of $x^{p^n} - x$. Now assuming that $g(x) \nmid f(x)$, since $g(x)$ is irreducible the GCD is 1, and so there exists $a(x), b(x)$ s.t.

$$a(x)f(x) + b(x)g(x) = 1 \tag{680}$$

But by setting $x = \alpha$, we get $f(\alpha) = 0$ from above, and $g(\alpha) = 0$ by assumption, leading to $0 = 1$, which is a contradiction since $0 \neq 1$ always in fields. Therefore $g(x) \mid f(x)$.

For (b), we have shown in (a) that the irreducible factors of $f(x)$ are precisely all polynomials in $\mathbb{Z}_p[x]$ whose degree divides n. Since $\mathbb{Z}_p$ is a field, we can scalar multiply the polynomial—and hence the leading coefficient—by the multiplicative inverse of the leading coefficient to make it monic. This doesn't change the factorization since the leading coefficient of $f(x)$ is also 1. Since $\mathbb{Z}_p[x]$ is a Euclidean domain, by unique factorization theorem all such polynomials $g(x)$ must be contained within the product.
It now remains to show that $f(x)$ is square free, i.e. none of its factors have multiplicity greater than 1. Take f and its derivative (where $p = 0$ in $\mathbb{Z}_p$)

$$f(x) = x^{p^n} - x, \qquad f'(x) = p^n x^{p^n - 1} - 1 = -1 \tag{681}$$

It is clear that $\gcd(f, f') = 1$ since f' is constant. Now assume that there is some factor $a(x)$ of multiplicity at least 2. Then $f(x) = a(x)a(x)b(x)$ for some $b(x) \in \mathbb{Z}_p[x]$. Taking the derivative gives

$$f'(x) = (a(x)a'(x) + a'(x)a(x))b(x) + a(x)^2 b'(x) \tag{682}$$

$$= a(x)\big(a'(x)b(x) + a'(x)b(x) + a(x)b'(x)\big) \tag{683}$$

which means that at least $a(x) \mid \gcd(f, f')$, contradicting that the gcd is 1. Therefore f is square free. Finally, since $\mathbb{Z}_p[x]$ is a Euclidean domain, by the unique factorization theorem all of its factors are precisely

Exercise 11.5 (Shifrin 7.6.2)

Let $F \supset \mathbb{Q}$ be a field extension. Prove that if $f(x) \in F[x]$ is irreducible, then it has no repeated roots in any field extension of F. (Hint: By Exercise 3.1.15, a repeated root must be a root of both $f(x)$ and its derivative.)

Solution. Assume that it is irreducible and for contradiction, let $f(x)$ have a repeated root α in some field extension $K \supset F$. Since α is a repeated root, we can write $f(x) = (x - \alpha)^2 g(x)$ for some $g(x) \in K[x]$. Then calculating its derivative we get

$$f'(x) = (x - \alpha)\big(2g(x) + (x - \alpha)g'(x)\big) \tag{684}$$

and so α is a root of both f and f', $\gcd(f(x), f'(x))$ has degree at least 1. This means that $f(x), f'(x)$ share a common factor in $F[x]$. Since $f(x)$ was irreducible over F, any nontrivial factor of $f(x)$ in $F[x]$ must be $f(x)$ itself. Therefore $f(x) \mid f'(x)$. However, this cannot be since $\deg(f'(x)) < \deg(f(x))$ when

$f(x)$ is a non-constant polynomial (which is true by definition since $f(x)$ is irreducible). Therefore, $f(x)$ cannot have a repeated root.

Exercise 11.6 (Shifrin 7.6.10)

Suppose $F \subset E \subset K$ are fields. Prove or give a counterexample:
a. $|G(K/F)| \leq |G(K/E)||G(E/F)|$
b. $|G(K/F)| \geq |G(K/E)||G(E/F)|$

Solution. (a) is false, and for a counterexample, take $F = \mathbb{Q} \subset E = \mathbb{Q}(\sqrt[3]{2}) \subset \mathbb{Q}(\sqrt[3]{2}, \omega)$ where ω is a cube root of unity not equal to 1. Then
1. $G(K/F)$ has 6 elements as shown in Shifrin Example 7.6.5.
2. $G(K/E)$ has 2 elements (the two automorphisms that fix $\sqrt[3]{2}$ and map ω to ω or ω^2).
3. $G(E/F)$ has 1 element as shown in Shifrin Example 7.6.5.

and so the inequality doesn't hold.

Exercise 11.7 (Shifrin 7.6.11)

Let $\mathbb{Q} \subset E \subset K$ be fields. Give a proof or counterexample for each of the following:
a. If K is a Galois extension of $\mathbb{Q}$, then K is a Galois extension of E.
b. If K is a Galois extension of $\mathbb{Q}$, then E is a Galois extension of $\mathbb{Q}$.
c. If K is a Galois extension of E and E is a Galois extension of $\mathbb{Q}$, then K is a Galois extension of $\mathbb{Q}$.

Solution. Listed.
1. True by the fundamental theorem of Galois theory.
2. Not true. Consider $\mathbb{Q} \subset E = \mathbb{Q}(\sqrt[3]{2}) \subset K = \mathbb{Q}(\sqrt[3]{2}, \omega)$, where ω is a cube root of unity. K is the splitting field of $x^3 - 2$ over $\mathbb{Q}$ so K is Galois over $\mathbb{Q}$. Also note that from Shifrin Example 7.6.5 the field K for the first two solutions is indeed Galois over $\mathbb{Q}$. E is not Galois over $\mathbb{Q}$ as shown in Shifrin Example 7.6.5.
3. Not true since we can consider the Galois extensions $\mathbb{Q} \subset \mathbb{Q}(\sqrt{2})$, which was proved to be Galois in Shifrin, and also $\mathbb{Q}(\sqrt{2}) \subset \mathbb{Q}(\sqrt[4]{2})$, which we will prove is Galois. Indeed, $\mathbb{Q}(\sqrt[4]{2})$ is the splitting field of the polynomial $f(x) = x^2 - \sqrt{2} = (x + \sqrt[4]{2})(x - \sqrt[4]{2}) \in \mathbb{Q}(\sqrt{2})[x]$ which doesn't have repeated roots, and so from lecture (Theorem 3 in April 16 Notes) it is Galois. However, $\mathbb{Q} \subset \mathbb{Q}(\sqrt[4]{2})$ is not Galois. Note that $\mathbb{Q}(\sqrt[4]{2})$ is a $\mathbb{Q}$-vector space with basis $\{1, \sqrt[4]{2}, \sqrt[4]{4}, \sqrt[4]{8}\}$, and so every element can be written as $a + b\sqrt[4]{2} + c\sqrt[4]{4} + d\sqrt[4]{8}$, and a $\mathbb{Q}$-automorphism must act on it as
$$\phi(a + b\sqrt[4]{2} + c\sqrt[4]{4} + d\sqrt[4]{8}) = a + b\phi(\sqrt[4]{2}) + c\phi(\sqrt[4]{2})^2 + d\phi(\sqrt[4]{2})^3 \tag{685}$$
So ϕ is determined by the value of $\phi(\sqrt{2})$. Since $\sqrt[4]{2}$ is the root of an irreducible polynomial $f(x) = x^4 - 2 \in \mathbb{Q}[x]$, any $\phi \in G(\mathbb{Q}(\sqrt[4]{2})/\mathbb{Q})$ must take roots to roots. The roots of $f(x)$ is $\{\pm\sqrt[4]{2}, \pm i\sqrt[4]{2}\}$. Since two of them lie outside of $\mathbb{Q}(\sqrt[4]{2})$, the Galois group can have at most 2 elements, which is less than $[\mathbb{Q}[\sqrt[4]{2}] : \mathbb{Q}] = 4$. Indeed, the splitting field of $x^4 - 2$ is $\mathbb{Q}(\sqrt[4]{2}, i)$.

Exercise 11.8 (Shifrin 7.6.14)

Suppose $f(x) \in \mathbb{Q}[x]$ is an irreducible polynomial of degree n. Prove that the Galois group of $f(x)$ must act transitively on the set of n roots. Give an example to show this statement is false for reducible polynomials. (Hint: Cf. the proof of Proposition 6.6.)

Solution. Let $f(x) \in \mathbb{Q}[x]$ be irreducible of degree n with $r_1, \dots, r_n$ its roots in a splitting field K. Since this is over $\mathbb{Q}$, the roots must be distinct. We wish to show that for any two roots $r_i \neq r_j$, there exists an element $\sigma \in G = G(K/\mathbb{Q})$ (which is a subgroup of the symmetric group over the roots) such that $\sigma(r_i) = r_j$. Since $f(x)$ is irreducible over $\mathbb{Q}$, the minimal polynomial of each root r_i over $\mathbb{Q}$ is $f(x)$. For any root r_i, consider the field $\mathbb{Q}(r_i)$. Since r_i is a root of the irreducible polynomial $f(x)$ of degree n, we have $[\mathbb{Q}(r_i) : \mathbb{Q}] = n$. For any two roots r_i, r_j, we can define a $\mathbb{Q}$-isomorphism $\mathbb{Q}(r_i) \to \mathbb{Q}(r_j)$ by setting $\phi(r_i) = r_j$ and extending linearly. By the fundamental theorem of Galois theory, this isomorphism ϕ extends to an automorphism $\sigma \in G$ of the splitting field K such that $\sigma(r_i) = r_j$, and we have found our desired σ.
For a counterexample, consider

$$f(x) = (x^2 - 2)(x^2 - 3) \in \mathbb{Q}[x] \tag{686}$$

which is reducible. The roots are $\pm\sqrt{2}, \pm\sqrt{3}$ and so the splitting field is $K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$ which is a vector space of dimension 4. So the Galois group $G = G(K/\mathbb{Q})$ has order 4 that sends $(\sqrt{2}, \sqrt{3})$ to $(\sqrt{2}, \sqrt{3}), (\sqrt{2}, -\sqrt{3}), (-\sqrt{2}, \sqrt{3}), (-\sqrt{2}, -\sqrt{3})$. Since there is no map $\sqrt{2} \mapsto \sqrt{3}$, G does not act transitively. Also consider the fact that

$$2 = \phi(2) = \phi(\sqrt{2})^2 = \sqrt{3}^2 = 3 \tag{687}$$

which is absurd.

Exercise 11.9 (Shifrin 7.6.23)

Let p be prime. Prove that the Galois group of $f(x) = x^p - 1$ is cyclic. (Hint: Use Exercise 6.1.24.)

Solution. The roots of $x^p - 1$ are the distinct pth roots of unity $1, \omega, \dots, \omega^{p-1}$. Therefore the splitting field is $\mathbb{Q}[\omega]$, which is Galois. So, for any $\phi \in G(\mathbb{Q}(\omega)/\mathbb{Q})$, $\phi(1) = 1$ by definition and since it must map roots to roots, $\phi(\omega) = \omega^k$ for some $1 \leq k \leq p-1$. Now define the map from the Galois group to $(\mathbb{Z}_p^*, \times)$ by

$$\sigma : \phi \to k \pmod p, \text{ where } \phi(\omega) = \omega^k \tag{688}$$

The map is well defined and a group homomorphism since if $\phi_1(\omega) = \omega^{k_1}$ and $\phi_2(\omega) = \omega^{k_2}$, then

$$(\phi_1 \circ \phi_2)(\omega) = \phi_1(\omega^{k_2}) = (\omega^{k_1})^{k_2} = \omega^{k_1 k_2} \tag{689}$$

It is injective since if $\phi_1 \neq \phi_2$, then they must send to different powers. Finally it is surjective since for any $k \in \mathbb{Z}_p^*$, we can define $\phi_k(\omega) = \omega^k$ and extend linearly. Therefore, σ is an isomorphism, and the Galois group is cyclic.

12 Affine and Projective Spaces

12.1 Affine Spaces

Modeling the space of points as a vector space can be unsatisfactory for a number of reasons.

1. The origin 0 plays a special role, when it doesn't necessarily need to have one.
2. Certain notions, such as parallelism, are handled in an awkward manner.
3. The geometries of vector and affine spaces are intrinsically. That is,

$$\mathrm{GL}(V) \subset \mathrm{GA}(V) \tag{690}$$

In the ordinary Euclidean geometry, one can define the operation of the addition of a point and a vector. That is, the "sum" of a point p and a vector x is the endpoint of a vector that starts at p and equals x. We formalize it in the following definition.

Definition 12.1

Let V be a vector space over field $\mathbb{F}$. The **affine space associated to** V is a set S with an operation of addition $+ : S \times V \longrightarrow S$ satisfying
1. $p + (x + y) = (p + x) + y$ for $p \in S, x, y \in V$
2. $p + 0 = p$ where $p \in S$, 0 is the zero vector
3. For any $p, q \in S$, there exists a unique vector x such that $p + x = q$

Elements of the set S are called **points**. The vector in condition 3 is called the **vector connecting points** p **and** q, denoted $\overline{pq}$. The dimension of an affine space is defined as the dimension of the corresponding vector space.

The first condition implies that

$$\overline{pq} + \overline{qr} = \overline{pr} \text{ for all } p, q, r \in S \tag{691}$$

Every vector space V can be regarded as an affine one if we view vectors both as points and as points and define the operation of addition of a vector to a point as addition of vectors. Under this interpretation, the vector $\overline{pq}$ is the difference between the vectors p and q.

Definition 12.2

Conversely, if we fix a point o (the origin) in an affine space S, we can identify a point p with its **position vector** $\overline{op}$. Then, addition of a vector to a point just becomes the addition a vectors. This identification of points with vectors is called the **vectorization** of an affine space.

Definition 12.3

A point o (the origin) together with a basis $\{e_1, ..., e_n\}$ of the space V is called a **frame** of the affine space S. Each frame is related to an **affine system of coordinates** in the space S. That is, a point p would get the coordinates equal to those of the vector $\overline{op}$ in the basis $\{e_1, ..., e_n\}$. It is easy to see that
1. Coordinates of the point $p + x$ are equal to the sums of respective coordinates of the point p and the vector x.
2. Coordinates of the vector $\overline{pq}$ are equal to the differences of respective coordinates of the points q and p.

Linear combinations of points are not defined in the affine space since the values of linear combinations are actually dependent on the choice of the origin. However, an analogous structure can be.

Definition 12.4

The **barycentric linear combination** of points $p_1, ..., p_k \in S$ is a linear combination of the form

$$p = \sum_i \lambda_i p_i, \text{ where } \sum_i \lambda_i = 1 \tag{692}$$

This linear combination is equal to the point p such that

$$\overline{op} = \sum_i \lambda_i \overline{op_i} \tag{693}$$

where $o \in S$ is any origin point.

Definition 12.5

In particular, the specific barycentric combination of points where $\lambda_1 = ... = \lambda_k = \frac{1}{k}$ is called the **center of mass** of the collection of points p_i.

Definition 12.6

Let $p_0, p_1, ..., p_n$ be points of an n-dimensional affine space S such that the vectors $\overline{p_0p_1}, ..., \overline{p_0p_n}$ are linearly independent (that is, forms a basis). Then, every point $p \in S$ can be uniquely presented as

$$p = \sum_{i=0}^{n} x_i p_i, \text{ where } \sum_{i=0}^{n} x_i = 1 \tag{694}$$

This equality can be rewritten

$$\overline{p_0p} = \sum_{i=1}^{n} x_i \overline{p_0p_i} \tag{695}$$

implying that we can take the coordinates of the vector $\overline{p_0p}$ in the basis $\{\overline{p_0p_1}, ..., \overline{p_0p_n}\}$ as $x_1, ..., x_n$. Then, x_0 is determined as

$$x_0 = 1 - \sum_{i=1}^{n} x_i \tag{696}$$

The numbers $x_0, x_1, ..., x_n$ are called the **barycentric coordinates** of the point p with respect to $p_0, p_1, ..., p_n$.

Definition 12.7

A **plane** in an affine space S is a subset of the form

$$p = p_0 + U \tag{697}$$

where p_0 is a point and U is a subspace of the space V. Note that we can choose any point p_0 in the plane in this representation. U is called the **direction subspace** for P.

Lemma 12.1

If the intersection of two planes in an affine space is nonempty, then the intersection is also a plane.

Theorem 12.2

Given any $k+1$ points of an affine space, there is a plane of dimension $\leq k$ passing through these points. If these points are not contained in a plane of dimension $< k$, then there exists a unique k-dimensional plane passing through them.

Definition 12.8

Points $p_0, p_1, ..., p_k \in S$ are **affinely dependent** if they lie in a plane of dimension $< k$, and **affinely independent** otherwise. It is clear that the points $p_0, ..., p_k$ are affinely independent if and only if the vectors $\overline{p_0p_1}, ..., \overline{p_0p_k}$ are linearly independent.

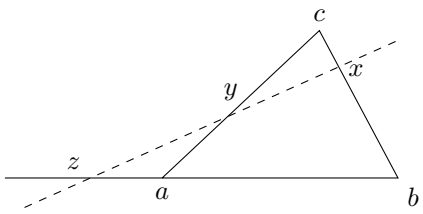
Theorem 12.3

Points $p_0, ..., p_k \in S$ are affinely independent if and only if the rank of the matrix of their barycentric coordinates (with respect to some predetermined affinely independent points) equals $k + 1$.

It is easy to see that the previous theorem is true, since the determinant represents the hypervolume of the parallelopiped spanned by the vectors $\overline{p_0p_1}, ..., \overline{p_0p_k}$, which must be nonzero if they are indeed affinely independent.

Corollary 12.4 (Menelaus' Theorem)

Let points x, y, z line on the sides bc, ca, ab of the triangle abc or their continuations.



Suppose that they divide these sides in the ratio

$$\lambda : 1, \mu : 1, \nu : 1$$

respectively. Then, the points x, y, z lie on the same line if and only if

$$\lambda\mu\nu = -1$$

Proof. By the previous theorem, the points x, y, z are linearly dependent (i.e. lies on a line) if and only if the matrix of barycentric coordinates of x, y, z with respect to a, b, c, which is

$$\begin{pmatrix} 0 & \frac{1}{\lambda+1} & \frac{\lambda}{\lambda+1} \\ \frac{\mu}{\mu+1} & 0 & \frac{1}{\mu+1} \\ \frac{1}{\nu+1} & \frac{\nu}{\nu+1} & 0 \end{pmatrix} \tag{698}$$

has nonzero determinant. The determinant of the above matrix is 0 if and only if $\lambda\mu\nu = -1$.

Corollary 12.5 (Ceva's Theorem)

In the triangle above, the lines ax, by, cz intersect at one point if and only if

$$\lambda\mu\nu = 1 \tag{699}$$

Proof. The proof can be done using barycentric coordinates.

Theorem 12.6

A nonempty subset $P \subset S$ is a plane if and only if for any two distinct points $a, b \in P$, the line through a and b also lies in P.

Theorem 12.7

Given an inhomogeneous system of linear equations of form

$$Ax = b \tag{700}$$

the set of solutions is an affine plane of dimension $n - r$, where n is the number of variables and r is the rank of the matrix A. More precisely, given that the plane is in the form $P = p_0 + U$, p_0 is one solution and U is the set of vectors that satisfy the homogeneous system

$$Ax = 0 \tag{701}$$

Let us observe the relative position of two planes.

Theorem 12.8

Given two planes

$$P_1 = p_1 + U_1, P_2 = p_2 + U_2$$

P_1 and P_2 intersect if and only if

$$\overline{p_1 p_2} \subset U_1 + U_2 \tag{702}$$

where $U_1 + U_2$ is the set of all vectors of form $u_1 + u_2$, where $u_1 \in U_1, u_2 \in U_2$.

Now, consider the class of functions on an affine space corresponding to the class of linear functions on a vector space.

Definition 12.9

An **affine-linear** function on an affine space S is a function $f : S \longrightarrow \mathbb{F}$ such that

$$f(p + x) = f(p) + \alpha(x), \quad p \in S, x \in V \tag{703}$$

where α, called the **differential**, is a linear function on the vector space V. Let $o \in S$ be a fixed origin. By setting $p = o$, we can express an affine linear function in vectorized form as

$$f(x) = \alpha(x) + b, \quad b \in \mathbb{F} \tag{704}$$

where $b = f(o)$. This implies the following coordinate form of f.

$$f(x) = b + \sum_i a_i x_i \tag{705}$$

A particular case of affine-linear functions are constant functions, where the defining characteristic is the zero differential.

Theorem 12.9

Given that $\dim S = n$, affine-linear functions on S form a $(n+1)$-dimensional subspace on the space of all linear functions on S.

Theorem 12.10

Barycentric coordinates are affine-linear functions.

Theorem 12.11

Let f be an affine-linear function. Then

$$f\left(\sum_i \lambda_i p_i\right) = \sum_i \lambda_i f(p_i) \tag{706}$$

for any barycentric linear combination $\sum_i \lambda_i p_i$ of points $p_1, ..., p_k$.

Definition 12.10

An affine space associated with a Euclidean vector space is called a **Euclidean affine space**. The **distance** ρ between two points in a Euclidean space is defined as

$$\rho(p, q) = ||\overline{pq}|| \tag{707}$$

This definition of ρ satisfies the axioms of a metric space.

12.2 Convex Sets

Let S be an affine space over the field of real numbers and V, the associated vector space.

Definition 12.11

The **(closed) interval** connecting points $p, q \in S$ is the set

$$pq = \{\lambda p + (1-\lambda)q \mid 0 \leq \lambda \leq 1\} \tag{708}$$

Geometrically, we can think of this as the straight line segment connecting point p with point q.

Definition 12.12

A set $M \subset S$ is **convex** if for any two points $p, q \in S$, it contains the whole interval p, q.

Clearly, the intersection of convex sets is convex. However, the union of them is not.

Definition 12.13

A **convex linear combination** of points in S is their barycentric linear combination with nonnegative coefficients.

It is clear to visualize the following theorem.

Theorem 12.12

For any points $p_0, ..., p_k$ in a convex set $M \subset S$, the set M also contains every convex linear combination

$$p = \sum_i \lambda_i p_i \tag{709}$$

Furthermore, for any set $M \subset S$, the set $\operatorname{conv} M$ of all convex linear combinations of points in M is convex.

Definition 12.14

Given $M \subset S$, the set $\operatorname{conv} M$ is the smallest convex set containing M. It is called the **convex hull** of M.

Definition 12.15

The convex hull of a system of affinely independent points $p_0, p_1, ..., p_n$ in an n-dimensional affine space is called the **n-dimensional simplex** with vertices $p_0, ..., p_n$.

It is clear that the interior points of a simplex is precisely the set of all points whose barycentric coordinates with respect to the vertices are all positive.

Example 12.1

Here are common examples of simplices.
1. A 0-dimensional simplex is a point.
2. A 1-dimensional simplex is a closed line interval.
3. A 2-dimensional simplex is a triangle.
4. A 3-dimensional simplex is a tetrahedron.

Theorem 12.13

A convex set M has interior points if and only if $\operatorname{aff} M = S$.

Definition 12.16

A convex set that has interior points is called a **convex body**. Clearly, every convex body in n-dimensional affine space S is n-dimensional.

The set of interior points of a convex body M, denoted M°, is an open convex body.

Definition 12.17

For any nonconstant affine-linear function f on the set S, let

$$H_f \equiv \{p \in S \mid f(p) = 0\}$$
$$H_f^+ \equiv \{p \in S \mid f(p) \geq 0\}$$
$$H_f^- \equiv \{p \in S \mid f(p) \leq 0\}$$

The set H_f is a hyperplane, and H_f^+, H_f^- are called **closed half spaces**.

Definition 12.18

A hyperplane H_f is a **supporting hyperplane** of a closed convex body M if $M \subset H_f^+$ and H_f contains at least one (boundary) point of M. The half space H_f^+ is then called the **supporting half-space** of M.

Theorem 12.14

A hyperplane H that passes through a boundary point of a closed convex body M, is supporting if and only if $H \cap M^\circ = \emptyset$.

A key theorem of convex sets is the following separation theorem.

Theorem 12.15 (Separation Theorem)

For every boundary point of a closed convex body, there exists a supporting hyperplane passing through this point.

This theorem leads to the following one.

Theorem 12.16

Every closed convex set M is an intersection of (perhaps infinitely many) half-spaces.

Definition 12.19

A **polyhedron** is the intersection of a finite number of half-spaces. A convex polyhedron which is also a body is called a **convex solid**.

Example 12.2

A simplex with vertices $p_0, p_1, ..., p_n$ is a convex polyhedron since it is determined by linear inequalities $x_i \geq 0$ for $i = 0, 1, ..., n$, where $x_0, x_1, ..., x_n$ are barycentric coordiantes with respect to $p_0, p_1, ..., p_n$.

Example 12.3

A convex polyhedron determined by linear inequalities $0 \leq x_i \leq 1$ for $i = 1, ..., n$, where $x_1, ..., x_n$ are affine coordinates with respect ot some frame, is called an n-dimensional parallelopiped.

Definition 12.20

A point p of a convex set M is **extreme** if it is not an interior point of any interval in M.

Theorem 12.17

A bounded closed convex set M is the convex hull of the set $E(M)$ of its extreme points.

We can create a stronger statement with the following theorem.

Theorem 12.18 (Minkowski-Weyl Theorem)

The following properties of a bounded set $M \subset S$ is equivalent.
1. M is a convex polyhedron.
2. M is a convex hull of a finite number of points.

Definition 12.21

A **face** of a convex polyhedron M is a nonempty intersection of M with some of its supporting hyperplanes. Given that $\dim \text{aff } M = n$,
1. A 0-dimensional face is called a **vertex**.
2. A 1-dimensional face an **edge**.
3. ...
4. An $(n-1)$-dimensional face a **hyperface**.

Therefore, if a convex polyhedron is determined by a system of linear inequalities, we can obtain its faces by replacing some of these inequalities with equalities (in such a way that we do not get the empty set).

The following theorem demonstrates that in order to find its faces, it suffices to consider only the hyperplanes $H_{f_1}, ..., H_{f_m}$.

Theorem 12.19

Every face Γ of the polyhedron M is of the form

$$\Gamma = M \cap \left(\bigcap_{j \in J} H_{f_j} \right) \tag{710}$$

where $J = \{1, 2, ..., m\}$

Theorem 12.20

The extreme points of a convex polyhedron M are exactly its vertices.

The following theorem is used often in linear programming and in optimization.

Theorem 12.21

The maximum of an affine-linear function on a bounded convex polyhedron M is attained at a vertex.

12.3 Affine Transformations and Motions

Let S and S' be affine spaces associated with vector spaces V and V', respectively, over the same field $\mathbb{F}$.

Definition 12.22

An **affine map** from the space S to the space S' is a map $f : S \longrightarrow S'$ such that

$$f(p + x) = f(p) + \varphi(x), \quad p \in S, x \in V \tag{711}$$

for some linear map $\varphi : V \longrightarrow V'$. It follows that

$$\varphi(\overline{pq}) = \overline{f(p)f(q)}, \quad p, q \in S \tag{712}$$

Thus, f determines the linear map φ uniquely. Similarly, φ is called the **differential** of f, denoted df.

Theorem 12.22

Let $f : S \longrightarrow S'$ and $g : S' \longrightarrow S''$ be two affine maps. Then the map

$$g \circ f : S \longrightarrow S'' \tag{713}$$

is also affine. Also

$$d(g \circ f) = dg \cdot df \tag{714}$$

where dg and df are the differentials of g and f, respectively.

For $\mathbb{F} = \mathbb{R}$, the differential of an affine map is a particular case of a differential of a smooth map in analysis. That is, the differential is the linear approximation of the function f.

Theorem 12.23

An affine map is bijective if and only if its differential is bijective.

Definition 12.23

Similar to linear transformations between vector spaces, bijective affine transformations are called **isomorphisms** of affine spaces. Affine spaces are **isomorphic** if there exists an isomorphism between them.

Corollary 12.24

Finite-dimensional affine spaces over the same field are isomorphic if and only if they have the same dimension.

Definition 12.24

An affine map from an affine space S to itself is called an **affine transformation**. Bijective affine transformations form a group called the **affine group of** S, denoted $\mathrm{GA}(S)$.

It follows that given affine space S with associated vector space V, the projection map

$$d : \mathrm{GA}(S) \longrightarrow \mathrm{GL}(V) \tag{715}$$

is a group homomorphism. It's kernel is the group of parallel translations, called $\mathrm{Tran}(S)$.

$$t_a : p \mapsto p + a, \quad a \in V \tag{716}$$

Theorem 12.25

For any $f \in \mathrm{GA}(S)$ and $a \in V$,

$$f t_a f^{-1} = t_{df(a)} \tag{717}$$

Definition 12.25

A **homothety** with the center o and coefficient λ is an affine transformation defined as

$$f(o + x) \equiv o + \lambda x \tag{718}$$

In its vectorized form, it is expressed

$$f(x) = \lambda x + b, \quad b \in V \tag{719}$$

A homothety with coefficient -1 is called a **central symmetry**.

The group of affine transformations determines the **affine geometry** of the space. The following theorem shows that all simplices are equal in affine geometry.

Theorem 12.26

Let $\{p_0, ..., p_n\}$ and $\{q_0, ..., q_n\}$ be two systems of affinely independent points in an n-dimensional affine space S. Then there exists a unique affine transformation f that maps p_i to q_i for $i = 0, 1, ..., n$.

Proof. It is easy to see once we realize that there exists a unique linear map φ of the space V that maps the basis $\{\overline{p_0 p_1}, ..., \overline{p_0 p_n}\}$ to the basis $\{\overline{q_0 q_1}, ..., \overline{q_0 q_n}\}$. If we vectorize S by taking p_0 as the origin, the affine transformation in question has the form

$$f(x) = \varphi(x) + \overline{p_0 q_0} \tag{720}$$

Corollary 12.27

In real affine geometry all parallelopipeds are equal.

Definition 12.26

A **motion** of the space S is an affine transformation of S whose differential is an orthogonal operator (i.e. an origin preserving isometry). Every motion is bijective.

Motions of a Euclidean space S form a group denoted $\mathrm{Isom}\, S$. A motion is called **proper (orientation preserving)** if its differential belongs to $\mathrm{SO}(V)$ and improper otherwise.

Lemma 12.28

The group $\mathrm{Isom}\, S$ is generated by reflections through hyperplanes.

Definition 12.27

Let M be a solid convex polyhedron in an n-dimensional Euclidean space. A **flag of** M is a collection of its faces $\{F_0, F_1, ..., F_{n-1}\}$ where $\dim F_k = k$ and $F_0 \subset F_1 \subset ... \subset F_{n-1}$.

Definition 12.28

A convex polyhedron M is **regular** if for any two of its flags, there exists a motion $f \in \operatorname{Sym} M$ mapping the first to the second, where

$$\operatorname{Sym} M \equiv \{f \in \operatorname{Isom} S \mid f(M) = M\} \tag{721}$$

Two dimensional regular polyhedra are the ordinary **regular polygons**. Their symmetry groups are known as the dihedral groups.

Three dimensional regular polyhedra are **Platonic solids**, which are the regular tetrahedron, cube, octahedron, dodecahedron, and icosahedron.

Definition 12.29

A real vector space V with a fixed symmetric bilinear function α of signature (k, l), where $k, l > 0$ and $\dim V = k + l$, is called the **pseudo-Euclidean vector space** of signature (k, l). The group of α-preserving linear transformations of V is called the **pseudo-orthogonal group** and is denoted $\mathrm{O}(V, \alpha)$. In an orthonormal basis, the corresponding matrix group is denoted Ok, l.

12.4 Quadrics

Planes are the simplest objects of affine and Euclidean geometry, which are determined by systems of linear equations. The second simplest are quadratic functions. These types of objects are studied futher in algebraic geometry.

Definition 12.30

An **affine-quadratic function** on an affine space S is a function $Q : S \longrightarrow \mathbb{F}$ such that its vectorized form is

$$Q(x) = q(x) + l(x) + c \tag{722}$$

for a quadratic function q, linear function l, and constant c.

12.5 Projective Spaces

Definition 12.31

An n-dimensional **projective space** PV over a field $\mathbb{F}$ is the set of one-dimensional subspaces of an $(n+1)$-dimensional vector space V over $\mathbb{F}$. For every $(k+1)$-dimensional subspace $U \subset V$, the subset $PU \subset PV$ is called a k-dimensional **plane** of the space PV.

1. 0-dimensional planes are the points of PV.
2. 1-dimensional planes are called **lines**
3. ...
4. $(n-1)$-dimensional planes are called **hyperplanes**

Definition 12.32

$\mathbb{RP}^1$ is called the real projective line, which is topologically equivalent to a circle.

Example 12.4

The real projective space of $\mathbb{R}^2$ is the set of all lines that pass through the origin. It is denoted $\mathbb{RP}^2$ and called the **real projective plane**.

Example 12.5

$\mathbb{RP}^3$ is diffeomorphic to SO(3).

Example 12.6

The space $\mathbb{RP}^n$ is formed by taking the quotient of $\mathbb{R}^{n+1} \setminus \{0\}$ under the equivalence relation

$$x \sim \lambda x \text{ for all real numbers } \lambda \neq 0 \tag{723}$$

The set of these equivalence classes is isomorphic to $\mathbb{RP}^n$.

12.6 Exercises

13 Representations

We will assume that V is a finite-dimensional vector space over field $\mathbb{C}$.

Definition 13.1

The **general linear group** of vector space V, denoted $\mathrm{GL}(V)$, is the group of all automorphisms of V to itself. The **special linear group** of vector space V, denoted $\mathrm{SL}(V)$ is the subgroup of automorphisms of V with determinant 1.

When studying an abstract set, it is often useful to consider the set of all maps from this abstract set to a well known set (e.g. $\mathrm{GL}(V)$).

Definition 13.2

A **representation** of an (algebraic) group $\mathcal{G}$ is a homomorphism

$$\rho : G \longrightarrow \mathrm{GL}(V) \tag{724}$$

for some vector space V. That is, given an element $g \in \mathcal{G}$, $\rho(g) \in \mathrm{GL}(V)$, meaning that $\rho(g)(v) \in V$. Additionally, since it is a homomorphism, the algebraic structure is preserved.

$$\rho(g_1 \cdot g_2) = \rho(g_1) \cdot \rho(g_2) \tag{725}$$

where $\cdot$ on the left hand side is the abstract group multiplication while the $\cdot$ on the right hand side is matrix multiplication. To shorten the notation, we will denote

$$gv = \rho(g)v, \ v \in V \tag{726}$$

Since ρ is a group morphism, we have

$$g_2(g_1 v) = (g_2 g_1)v \iff \rho(g_2)\big(\rho(g_1)(v)\big) = \big(\rho(g_2)\rho(g_1)\big)(v) \tag{727}$$

Additionally, since g (that is, $\rho(g)$) is a linear map,

$$g(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 g v_1 + \lambda_2 g v_2 \tag{728}$$

Usually, we refer to the map as the representation, but if the map is well-understood, we just call the vector space V the representation and say that the group acts on this vector space.

Example 13.1

The group $\mathrm{GL}(2, \mathbb{C})$ can be represented a by the vector space $\mathbb{C}^2$, or explicitly, by the group of 2×2 matrices over $\mathbb{C}$ with nonzero determinant.

$$\mathrm{GL}(2, \mathbb{C}) \overset{id}{\longmapsto} \mathrm{Mat}(2, \mathbb{C}) \tag{729}$$

This is a trivial representation.

We now show a nontrivial representation of $\mathrm{GL}(2, \mathbb{C})$.

Example 13.2

We take $\text{Sym}^2\mathbb{C}^2$, the second symmetric power of $\mathbb{C}^2$. Note that given a basis $x_1, x_2 \in \mathbb{C}^2$, the set

$$\{x_1 \odot x_1, x_1 \odot x_2, x_2 \odot x_2\} \tag{730}$$

forms a basis of $\text{Sym}^2\mathbb{C}^2 \implies \dim \text{Sym}^2\mathbb{C}^2 = 3$. So, we want to represent $\text{GL}(2, \mathbb{C})$ by associating its element with elements of $\text{GL}(Sym^2\mathbb{C}^2)$. More concretely, we are choosing to represent a 2×2 matrix over $\mathbb{C}$ with a 3×3 matrix group (since $\text{GL}(Sym^2\mathbb{C}^2) \simeq \text{GL}(3, \mathbb{C})$. Clearly,

$$\begin{aligned}
\rho(g)(x_1 \odot x_1) &= g(x_1) \odot g(x_1) \in Sym^2\mathbb{C}^2 \\
\rho(g)(x_1 \odot x_2) &= g(x_1) \odot g(x_2) \\
\rho(g)(x_2 \odot x_2) &= g(x_2) \odot g(x_2)
\end{aligned}$$

To present this in matrix form, let us have an element in $\text{GL}(2, \mathbb{C})$

$$\mathcal{A} \equiv \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{731}$$

We evaluate the corresponding representation in $\text{GL}(Sym^2\mathbb{C}^2)$. Using the identities above, we have

$$\begin{aligned}
\rho(g)(x_1 \odot x_1) &= g(x_1) \odot g(x_1) \\
&= (ax_1 + cx_2) \odot (ax_1 + cx_2) \\
&= a^2 x_1 \odot x_1 + 2acx_1 \odot x_2 + c^2 x_2 \odot x_2 \\
\rho(g)(x_1 \odot x_2) &= g(x_1) \odot g(x_2) \\
&= (ax_1 + cx_2) \odot (bx_1 + dx_2) \\
&= abx_1 \odot x_1 + (ad + bc)x_1 \odot x_2 + cdx_2 \odot x_2 \\
\rho(g)(x_2 \odot x_2) &= g(x_2) \odot g(x_2) \\
&= (bx_1 + dx_2) \odot (bx_1 + dx_2) \\
&= b^2 x_1 \odot x_1 + 2bdx_1 \odot x_2 + d^2 x_2 \odot x_2
\end{aligned}$$

And this completely determines the matrix. So,

$$\rho \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a^2 & ab & b^2 \\ 2ac & ad + bc & 2bd \\ c^2 & cd & d^2 \end{pmatrix} \tag{732}$$

is the 3×3 representation of $\mathcal{A}$ in $\text{GL}(Sym^2\mathbb{C}^2)$.

We continue to define maps between two representations of $\mathcal{G}$.

Definition 13.3

A **morphism** between 2 representations

$$\begin{aligned}
\rho_1 &: \mathcal{G} \longrightarrow \text{GL}(V_1) \\
\rho_2 &: \mathcal{G} \longrightarrow \text{GL}(V_2)
\end{aligned}$$

of some group but not necessarily the same vector space is a linear map $f : V_1 \longrightarrow V_2$ that is **compatible** with the group action. That is, f satisfies the property that for all $g \in \mathcal{G}$

$$f \circ g = g \circ f \tag{733}$$

Again, we use the shorthand notation that $g = \rho(g)$, meaning that the statement above really translates to $f \circ \rho(g) = \rho(g) \circ f$. This is equivalent to saying that the following diagram commutes.

$$\begin{array}{ccc} V_1 & \xrightarrow{\rho_1(g)} & V_1 \\ \downarrow f & & \downarrow f \\ V_2 & \xrightarrow{\rho_2(g)} & V_2 \end{array}$$

Definition 13.4

Let V be a representation of $\mathcal{G}$. A **subrepresentation** is a subspace $W \subset V$ such that for all $g \in \mathcal{G}$ and for all $w \in W$,

$$\rho(g)(w) \in W \tag{734}$$

Example 13.3

V and $\{0\}$ are always subrepresentations of V.

We now introduce the "building blocks" of all representations.

Definition 13.5

A representation W is **irreducible representation** if $\{0\}$ and W are the only subrepresentations of W.

Lemma 13.1 (Schur's Lemma)

Let V_1, V_2 be irreducible representations and let $f : V_1 \longrightarrow V_2$ be a morphism (of representations). Then, either

1. f is an isomorphism.
2. $f = 0$

Furthermore, any 2 isomorphisms differ by a constant. That is,

$$f_1 = \lambda f_2 \tag{735}$$

Proof. $\ker f$ is clearly a vector space. Furthermore, it is a subrepresentation (since it is a subspace of V_1) $\implies$ $\ker f = V$ or $\ker f = 0$. If $\ker f = V$, then $f = 0$ and the theorem is satisfied. If $\ker f = 0$, then f is injective, and $\text{Im}\, f$ is a subrepresentation of V_2 $\implies$ $\text{Im}\, f = 0$ or $\text{Im}\, f = V_2$. But $\text{Im}\, f \neq 0$ since f is injective, so $\text{Im}\, f = V_2$ $\implies$ f is surjective $\implies$ f is bijective, that is, f is an isomorphism of vector spaces. So, the inverse f^{-1} exists, and this map f^{-1} satisfies

$$f^{-1} \circ \rho_2(g) = \rho_1(g) \circ f^{-1} \tag{736}$$

To prove the second part, without loss of generality, assume that the first isomorphism is the identity mapping. That is,

$$f_1 = id \tag{737}$$

Since we are working over the field $\mathbb{C}$, we can find an eigenvector of f_2. That is, there exists a $v \in V_1$ such that

$$f_2(v) = \lambda v \tag{738}$$

Now, we define the map

$$f : V_1 \longrightarrow V_2,\ f \equiv f_2 - \lambda f_1 \tag{739}$$

Clearly, $\ker f \neq 0$, since $v \in \ker f$. That is, we have a map f between 2 irreducible representations that has a nontrivial kernel. This means that $f = 0 \implies f_2 = \lambda f_1$.

Theorem 13.2 (Mache's Theorem)

Let V be finite dimensional, with $\mathcal{G}$ a finite group. Then, V can be decomposed as

$$V = \bigoplus_i V_i \tag{740}$$

where each V_i is an irreducible representation of $\mathcal{G}$.

Proof. By induction on dimension, it suffices to prove that if W is a subrepresentation of V, then there exists a subrepresentation $W' \subset V$ such that $W \oplus W' = V$. So, if V isn't an irreducible representation, it can always be decomposed into smaller subrepresentations W and W' that direct sum to V. Now, we define the canonical (linear) projection

$$\pi : V \longrightarrow W \tag{741}$$

Then, we define the new map

$$\tilde{\pi} : V \longrightarrow W,\ \tilde{\pi}(v) \equiv \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} \rho(g)\big|_W \circ \pi \circ \rho(g)^{-1} \tag{742}$$

This "averaging" of the group elements are done so that this mapping is a map of representations. This implies that

$$V = W \oplus \ker \tilde{\pi} \tag{743}$$

meaning that V can indeed be decomposed into direct sums of subrepresentations.

13.1 Exercises

14 Lie Groups and Lie Algebras

Definition 14.1

A **Lie group** is a group $\mathcal{G}$ that is also a finite-dimensional smooth manifold, in which the group operations of multiplication and inversion are smooth maps. Smoothness of the group multiplication

$$\mu : \mathcal{G} \times \mathcal{G} \to \mathcal{G},\ \mu(x, y) = xy \tag{744}$$

means that μ is a smooth mapping of the product manifold $\mathcal{G} \times \mathcal{G}$ into $\mathcal{G}$. These two requirements can be combined to the single requirement that tahe mapping

$$(x, y) \mapsto x^{-1}y \tag{745}$$

be a smooth mapping of the product manifold into $\mathcal{G}$.

Definition 14.2

A **Lie Algebra** is a vector space $\mathfrak{g}$ with an operation called the **Lie Bracket**

$$[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g} \tag{746}$$

Satisfying
1. Bilinearity: $[ax + by, z] = a[x, z] + b[y, z],\ [z, ax + by] = a[z, x] + b[z, y]$
2. Anticommutativity: $[x, y] = -[y, x]$
3. Jacobi Identity: $[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0$

Clearly, this implies that $\mathfrak{g}$ is a nonassociative algebra. Note that a Lie Algebra does not necessarily need to be an algebra in the sense that there needs to be multiplication operation that is closed in $\mathfrak{g}$.

Example 14.1

A common example of a Lie Braket in the algebra of matrices is defined

$$[A, B] \equiv AB - BA \tag{747}$$

called the **commutator**. Note that in this case, the definition of the Lie bracket is dependent on the definition of the matrix multiplication. Without defining the multiplication operation, we wouldn't know what AB or BA means. Therefore, we see that the Lie algebra of $n \times n$ matrices has three operations: matrix addition, matrix multiplication, and the commutator (along with scalar multiplication). But in general, it is not necessary to have that multiplication operation for abstract Lie algebras. $\mathfrak{g}$ just needs to be a vector space with the bracket.

Example 14.2

The set of all symmetric matrices is a vector space, but it is **not** a Lie algebra since the commutator $[A, B]$ is not symmetric unless $AB = BA$.

We will first talk about groups of matrices as a more concerete example before we get into abstract Lie groups. Recall that the matrix exponential map is defined

$$exp : \text{Mat}(n, \mathbb{C}) \to \text{mat}(n, \mathbb{C}),\ exp(A) = e^A = \sum_{p \geq 0} \frac{A^p}{p!} \tag{748}$$

Note that this value is always well defined. This lets us define

$$exp(tA) \equiv e^{tA} \equiv I + tA + \frac{1}{2}t^2A^2 + \frac{1}{3!}t^3A^3 + ... \tag{749}$$

where if t is small, we can expect a convergence. Note that exp maps addition to multiplication. That is, we can interpret it as a homomorphism from

$$exp : \mathfrak{g} \to \mathcal{G} \tag{750}$$

where $\mathfrak{g}$ is the Lie algebra and $\mathcal{G}$ is the Lie group (which we will treat just as a matrix group). To find the inverse of the exponential map, we can take the derivative of e^{tA} at $t = 0$. That is,

$$\left(\frac{d}{dt}e^{tA}\right)\Big|_{t=0} = \left(\sum_{k=0}^{\infty}\frac{1}{k!}t^kA^{k+1}\right)\Big|_{t=0} = A$$

So, the mapping

$$\frac{d}{dt}\Big|_{t=0} : \mathcal{G} \to \mathfrak{g} \tag{751}$$

maps the Lie group back to the algebra. We can interpret this above mapping by visualizing the Lie Algebra as a tangent (vector) space of the abstract Lie group $\mathcal{G}$ at the identity element of the Lie group. The visualization below isn't the most abstract one, but it may help:

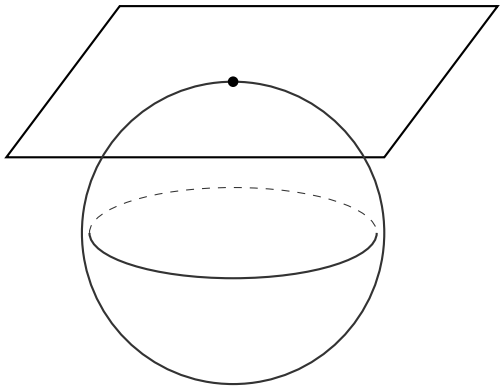
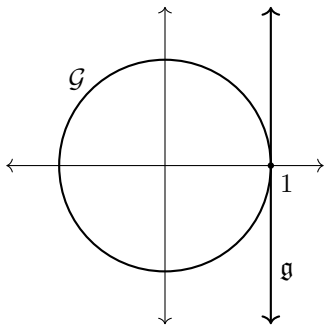


Figure 21: The Lie algebra can be visualized as the tangent space at the identity.

For example, say that the Lie group $\mathcal{G}$ is a unit circle in $\mathbb{C}$, then the Lie algebra of $\mathcal{G}$ is the tangent space at the identity 1, which can be identified as the imaginary line in the complex plane $\{it \mid t \in \mathbb{R}\}$, with

$$it \mapsto exp(it) \equiv e^{it} \equiv \cos t + i \sin t \tag{752}$$



So, analyzing the Lie group by looking at its Lie algebra turns a nonlinear problem to a linear one; this is called a **linearization** of the Lie group. The existence of this exponential map is one of the primary reasons that Lie algebras are useful for studying Lie groups.

Example 14.3

The exponential map

$$exp : \mathbb{R} \to \mathbb{R}^+,\ x \mapsto e^x \tag{753}$$

is a group homomorphism that maps $(\mathbb{R}, +)$ to $(\mathbb{R}^+, \times)$. This means that $\mathbb{R}$ is the Lie algebra of the Lie group $\mathbb{R}^+$.

Theorem 14.1

If A and B are commuting square matrices, then

$$e^{A+B} = e^A\, e^B \tag{754}$$

In general, the solution C to the equation

$$e^A\, e^B = e^C \tag{755}$$

is given by the **Baker-Campbell-Hausdorff formula**, defined

$$C = A + B + \frac{1}{2}[A, B] + \frac{1}{12}[A, [A, B]] - \frac{1}{12}[B, [A, B]] + ... \tag{756}$$

consisting of terms involving higher commutators of A and B. The full series is much too complicated to write, so we ask the reader to be satisfied with what is shown.

The BCH formula is messy, but it allows us to compute products in the Lie Group as long as we known the commutators in the Lie Algebra.

Therefore, we can describe the process of constructing a Lie group from a Lie Algebra (which a vector space) as such. We take a vector space V and endow it the additional bracket operation. We denote this as

$$\mathfrak{g} \equiv (V, [\cdot, \cdot]) \tag{757}$$

Then, we take every element of $\mathfrak{g}$ and apply the exponential map to them to get an another set $\mathcal{G}$. We then endow a group structure on $\mathcal{G}$ by defining the multiplication as

$$\cdot : \mathcal{G} \times \mathcal{G} \to \mathcal{G},\ e^A \cdot e^B = e^{A*B} \tag{758}$$

where $A * B$ is defined by the BCH formula up to a certain kth order. Since the $*$ operation is completely defined by the bracket in the Lie algebra, it tells us how to multiply in the Lie group. This process can be made more abstractly, depending on what A, B and $[\cdot, \cdot]$ is, beyond matrices.

14.1 Lie Algebras of Classical Lie Groups

Definition 14.3 (General Linear Group)

The **general linear group**, denoted $\mathrm{GL}(V)$, is the set of all bijective linear mappings from V to itself. Similarly, $\mathrm{GL}_n(\mathbb{F})$, or $\mathrm{GL}\ (n, \mathbb{F})$ is the set of all nonsingular $n \times n$ matrices over the field $\mathbb{F}$. Due to the same dimensionality of the following spaces, it is clear that $\mathrm{GL}(V) \simeq \mathrm{GL}(\mathbb{F}^n) \simeq \mathrm{GL}_n(\mathbb{F})$. The **special linear group**, denoted $\mathrm{SL}_n(\mathbb{F})$ or $\mathrm{SL}(n, \mathbb{F})$, is the set of $n \times n$ matrices a with determinant 1. $\mathrm{SL}_n(\mathbb{F})$ is a subgroup of $\mathrm{GL}_n(\mathbb{F})$, which is a subset of the ring of all $n \times n$ matrices over field $\mathbb{F}$, denoted $\mathbb{L}_n(\mathbb{F})$.

Definition 14.4 (Translation Group)

The group of all translations in the space V is denoted $\text{Tran}\, V$. Its elements are usually denoted as t_u, where u is the vector that is being translated by. It can also be interpreted as shifting the origin by $-u$. It is clear that $\text{Tran}\, V \simeq V$.

Definition 14.5 (General Affine Group)

The **general affine group** is the pair of all transformations

$$\text{GA}(V) \equiv \text{Tran}(V) \times \text{GL}(V) \tag{759}$$

Definition 14.6 (Isometries)

The **Euclidean group** of **isometries** in the Euclidean space $\mathbb{E}^n$ (with the Euclidean norm), denoted $\text{Isom}\, \mathbb{E}^n$ or $\mathbb{E}(n)$, consists of all distance-preserving bijections from $\mathbb{E}^n$ to itself, called **motions** or **rigid transformations**. It consists of all combinations of rotations, reflections, and translations. The **special Euclidean group** of all isometries that preserve the **handedness** of figures is denoted $\mathbb{SE}(n)$, which is comprised of all combinations rotations and translations called **rigid motions** or **proper rigid transformations**.

Definition 14.7 (Orthogonal Group)

The **orthogonal group**, denoted $\text{O}(n)$, consists of all isometries that preserve the origin, i.e. consists of rotations and reflections. The **special orthogonal group**, denoted $\text{SO}(n)$, is a subgroup of $\text{O}(n)$ consisting of only rotations. We can see that

$$\text{O}(n) = \frac{\text{Isom}\, \mathbb{E}^n}{\text{Tran}\, V} \tag{760}$$

Definition 14.8 (Transitive)

A transformation group G is called **transitive** if for any $x, y \in X$, there exists a $\phi \in G$ such that $y = \phi(x)$.

Example 14.4

$\text{Tran}(V)$ and $\text{GA}(V)$ are transitive groups.

Definition 14.9 (Congruence Classes)

Let X be a set and G its transformation group on X. The way we define G determines the **geometry** of X. More specifically, a figure $F_1 \subset X$ is **equivalent** or **congruent** to $F_2 \subset X$ iff there exists $\phi \in G$ such that $F_2 = \phi(F_1)$ (or equivalently, $F_1 = \phi(F_2)$). This is an equivalence relation since

1. $F \sim F$.
2. $F \sim H \implies H \sim F$.
3. $F \sim H, H \sim K \implies F \sim K$

Two figures that are in the same equivalence class are known to be **congruent** with respect to the geometry of X induced by G.

Clearly, if two figures are congruent in Euclidean geometry, then they are congruent in Affine geometry, since $\mathrm{E}(n) \subset \mathrm{GA}(n)$.

14.1.1 Lie Algebras of $\mathrm{SL}(2,\mathbb{R})$ and $\mathrm{SL}(2,\mathbb{C})$

Given the group $\mathrm{SL}(2,\mathbb{R})$, there must be a corresponding Lie algebra of matrices such that $g = e^A \in \mathrm{SL}(2,\mathbb{R})$. We attempt to find this Lie algebra. Let $g \in \mathrm{SL}(2,\mathbb{R})$, with $g = e^A$. So, if $\det g = 1$, what is the corresponding restriction on A in the algebra? We use the following theorem.

Theorem 14.2

$$\det\left(e^A\right) = e^{\mathrm{Tr}\,(A)} \tag{761}$$

Proof. Put A in Jordan Normal Form: $A = S^{-1}JS \implies A^n = S^{-1}J^nS \implies exp(A) = S^{-1}exp(A)S \implies \det\left(exp(A)\right) = \det e^J$. But since J is upper trianglar, J^n is upper triangular $\implies e^J$ is upper triangular, which implies that

$$\det e^J = \prod_i e^{\lambda_i} = e^{\mathrm{Tr}\,(J)} = e^{\mathrm{Tr}\,(A)} \tag{762}$$

since trace is invariant under a change of basis.

So, $\det\left(e^A\right) = 1 \implies \mathrm{Tr}\,(A) = 2\pi i n$ for $n \in \mathbb{Z}$. Since we want to component connected to the identity, we choose $n = 0$ meaning that $\mathrm{Tr}\,(A) = 0$. And we are done. That is, the Lie algebra of $\mathrm{SL}(2,\mathbb{R})$ consists of traceless 2×2 matrices, denoted $\mathfrak{sl}_2\mathbb{R}$. $\mathfrak{sl}_2\mathbb{R}$ has basis (chosen arbitrarily)

$$\left\{ H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, X = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, Y = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \right\} \tag{763}$$

and the identity in the Lie algebra is the zero matrix, which translates to the 2×2 identity matrix in the Lie group.

$$exp \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = I \tag{764}$$

We must not forget to define the bracket structure in $\mathfrak{sl}_2\mathbb{R}$, so we define it as the commutator, which gives the identity

$$\begin{aligned} [H, X] &= HX - XH = 2X \\ [H, Y] &= HY - YH = -2Y \\ [X, Y] &= XY - YX = H \end{aligned}$$

Note that regular matrix multiplication is not closed within this Lie algebra. For example,

$$XY = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{765}$$

is clearly not traceless. However, the bracket operation keeps the matrices within this traceless condition (and thus, within this algebra), so you can't just stupidly multiply matrices together in a Lie algebra. Remember that regular matrix multiplication does not have anything to do with the Lie bracket and does not apply to this group. This algebra also simplifies the multiplicative inverse of a group to a simple additive inverse, making calculations easier.

Similarly, the Lie algebra of $\mathrm{SL}(2,\mathbb{C})$ also has the same basis

$$\left\{ H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, X = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, Y = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \right\} \tag{766}$$

but we choose the field to be $\mathbb{C}$, meaning that we take complex linear combinations rather than real linear ones.

14.1.2 Lie Algebra of $\mathrm{SU}(2)$

$g \in \mathrm{SU}(2) \implies \det g = 1 \implies \mathrm{Tr}\, A = 0$. We also see that by definition e^A,

$$(e^A)^\dagger = e^{A^\dagger} \text{ and } (e^A)^{-1} = e^{-A} \tag{767}$$

which implies that $A^\dagger = -A$. That is, the unitary condition implies that the Lie algebra elements in $\mathfrak{su}(2)$ are traceless, anti-self adjoint 2×2 matrices over $\mathbb{C}$.

Definition 14.10

The **Pauli matrices** are the three matrices

$$\left\{ \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right\} \tag{768}$$

Note that with some calculation,

$$\begin{aligned} [\sigma_x, \sigma_y] &= 2i\sigma_z \\ [\sigma_y, \sigma_z] &= 2i\sigma_x \\ [\sigma_z, \sigma_x] &= 2i\sigma_y \end{aligned}$$

To identify the basis of $\mathfrak{su}(2)$, we take the Pauli matrices and let

$$\begin{aligned} A_x &\equiv -\frac{i}{2}\sigma_x = \begin{pmatrix} 0 & -i/2 \\ -i/2 & 0 \end{pmatrix} \\ A_y &\equiv -\frac{i}{2}\sigma_y = \begin{pmatrix} 0 & -1/2 \\ 1/2 & 0 \end{pmatrix} \\ A_z &\equiv -\frac{i}{2}\sigma_z = \begin{pmatrix} -i/2 & 0 \\ 0 & i/2 \end{pmatrix} \end{aligned}$$

be the basis of $\mathfrak{su}(2)$. Clearly, A_x, A_y, A_z are all traceless, anti-self adjoint 2×2 matrices. Moreover, they also satisfy

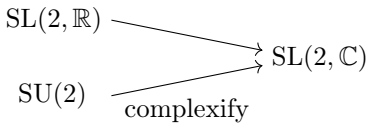
$$\begin{aligned} [A_x, A_y] &= A_z \\ [A_y, A_z] &= A_x \\ [A_z, A_x] &= A_y \end{aligned}$$

However, note that the algebra $\mathfrak{su}(2)$ consists of all **real** linear combinations of A_x, A_y, A_z. That is, $\mathfrak{su}(2)$ is a 3 dimensional **real** vector space, even though it has basis elements containing complex numbers.

However, we can always complexify this space by simply replacing real scalar multiplication in $\mathfrak{su}(2)$ with complex scalar multiplication. By complexifying $\mathfrak{su}(2)$, the Lie group $\mathrm{SU}(2)$ formed by taking the exponential map on this complexified space is actually identical to $\mathrm{SL}(2, \mathbb{C})$. Indeed, this is true because first, the basis $\{H, X, Y\}$ of $\mathfrak{sl}_2\mathbb{C}$ and the basis $\{A_x, A_y, A_z\}$ of $\mathfrak{su}(2)$ span precisely the same subspace in the vector space $\mathrm{Mat}(2, \mathbb{C})$, meaning that the two Lie algebras are the same vector space. Secondly, the bracket operation $[\cdot, \cdot]$ in both $\mathfrak{sl}_2\mathbb{C}$ and $\mathfrak{su}(2)$ are equivalent since the operation defined to be the commutator in both cases, resulting in the similarities in the bracket behaviors.

$$\begin{aligned} [H, X] = 2X &\iff [A_x, A_y] = A_z \\ [H, Y] = -2Y &\iff [A_y, A_z] = A_x \\ [X, Y] = H &\iff [A_z, A_x] = A_y \end{aligned}$$

Therefore, the complexification of $\mathrm{SU}(2)$ and $\mathrm{SL}(2, \mathbb{R})$ both leads to the construction of $\mathrm{SL}(2, \mathbb{C})$.



We can interpret the "real forms" of $\text{SL}(2,\mathbb{C})$ as "slices" of some complex group. However, this does not mean that the real version of these groups are equal. That is,

$$\text{SL}(2,\mathbb{R}) \neq \text{SU}(2) \tag{769}$$

14.1.3 Lie Algebra of SO(3)

It is easy to see that for SO(2), it is easy to see that its Lie algebra $\mathfrak{so}(2)$ has

$$\left\{ \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \right\} \tag{770}$$

as its only basis, since

$$exp\left(\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \theta \right) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{771}$$

meaning that the dimension of SO(2) is 1. By adding a component, we can get a rotation in $\mathbb{R}^3$.

$$\begin{aligned} R_x &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \implies e^{R_x} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \\ R_y &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} \implies e^{R_y} = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix} \\ R_z &= \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \implies e^{R_z} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

That is, e^{R_x}, e^{R_y}, and e^{R_z} generates a rotation around the x, y, and z axis, respectively, which completely generates the group SO(3). Therefore, the Lie algebra $\mathfrak{so}(3)$ consists of the basis

$$\{R_x, R_y, R_z\} \tag{772}$$

The bracket structure (again, defined as the commutator) of this Lie algebra is

$$\begin{aligned} [R_x, R_y] &= R_z \\ [R_y, R_z] &= R_x \\ [R_z, R_x] &= R_y \end{aligned}$$

which is similar to the brakcet structure of $\mathfrak{su}(2)$. Therefore, SO(3) and SU(2) have the **same** Lie algebra, which is the algebra of dimension 3 with the same bracket structure. Note that Lie algebras are uniquely determined by the bracket structure and dimension. However, having the same Lie algebra does not imply that the groups are identical (obviously) nor isomorphic. For example,

$$exp(2\pi R_z) = \begin{pmatrix} \cos 2\pi & -\sin 2\pi & 0 \\ \sin 2\pi & \cos 2\pi & 0 \\ 0 & 0 & 1 \end{pmatrix} = I \tag{773}$$

while

$$exp(2\pi A_z) = exp(-i\pi\sigma_z) = exp\left(-i\pi \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right) = -I \tag{774}$$

There is discrepancy by a factor of -1. In fact, it turns out that

$$\mathrm{SO}(3) = \frac{\mathrm{SU}(2)}{\pm I} \tag{775}$$

We justify this in the following way. Let $v \in \mathbb{R}^3$ have components (x, y, z). Consider

$$M = x\sigma_x + y\sigma_y + z\sigma_z \tag{776}$$

M is clearly traceless and $M^\dagger = M$. Now, let $S \in \mathrm{SU}(2)$ and let $M' = S^{-1}MS$. Then, $\mathrm{Tr}\, M' = \mathrm{Tr}\, S^{-1}MS = \mathrm{Tr}\, M = 0$ and $(M')^\dagger = (S^{-1}MS)^\dagger = S^\dagger M^\dagger (S^{-1})^\dagger = S^{-1}MS = M'$. Therefore, since M' is self adjoint and traceless, it can be expressed in the form

$$x'\sigma_x + y'\sigma_y + z'\sigma_z \tag{777}$$

for some (x', y', z'). Now, since

$$M^2 = (-x^2 - y^2 - z^2)I \tag{778}$$

we have

$$\begin{aligned}(M')^2 &= S^{-1}M^2S = (-x^2 - y^2 - z^2)I \\ &= (-x'^2 - y'^2 - z'^2)I\end{aligned}$$

So, $x^2 + y^2 + z^2 = x'^2 + y'^2 + z'^2$, implying that the lengths of v stayed the same. (The proof of linearity of S is easy.) Therefore, the transformation $M \mapsto M'$, i.e. $(x, y, z) \mapsto (x', y', z')$ is a linear transformation preserving length in $\mathbb{R}^3$ (with respect to the usual inner product and norm) $\implies$ it is in SO(3). If we have

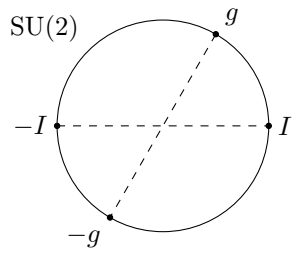
$$S = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \tag{779}$$

then $M' = M$, which explains why SO(3) is a coset deviating by both I and $-I$. Visually, if we let SU(2) be a circle, points that are diametrically opposite of each other are "equivalent" in SO(3). That is, SU(2) is a three-dimensional sphere, and g and $-g$ are identified onto the same element in SO(3). This map

$$\rho : \mathrm{SU}(2) \to \mathrm{SO}(3) \tag{780}$$

in which 2 points are mapped to 1 point is a surjective map with

$$\ker \rho = \{I, -I\} \tag{781}$$



We can in fact explicitly describe exponential map from $\mathfrak{so}(3)$ to SO(3) with the following lemma.

Lemma 14.3 (Rodrigues' Formula)

The exponential map $exp : \mathfrak{so}(3) \to \mathrm{SO}(3)$ is defined by

$$e^A = \cos\theta I_3 + \frac{\sin\theta}{\theta}A + \frac{(1 - \cos\theta)}{\theta^2}B \tag{782}$$

where

$$A = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix}, B = \begin{pmatrix} a^2 & ab & ac \\ ab & b^2 & bc \\ ac & bc & c^2 \end{pmatrix} \tag{783}$$

This formula has many applications in kinematics, robotics, and motion interpolation.

Theorem 14.4

The Lie algebras for the following classical Lie groups are summarized as follows.

1. $\mathfrak{sl}_n\mathbb{R}$ is the real vector space of real $n \times n$ matrices with null trace.
2. $\mathfrak{so}(n)$ is the real vector space of real $n \times n$ skew-symmetric matrices.
3. $\mathfrak{gl}_n\mathbb{R}$ is the real vector space of all real $n \times n$ matrices.
4. $\mathfrak{o}(n) = \mathfrak{o}(n)$

Note that the corresponding groups $\mathrm{GL}(n, \mathbb{R}), \mathrm{SL}(n, \mathbb{R}), \mathfrak{gl}_n\mathbb{R}, \mathfrak{sl}_n\mathbb{R}$ are Lie groups, meaning that they are smooth real manifolds. We can view each of them as smooth real manifolds embedded in the n^2 dimensional vector space of real matrices, which is isomorphic to $\mathbb{R}^{n^2}$.

Theorem 14.5

The Lie algebras $\mathfrak{gl}_{\mathbb{R}}, \mathfrak{sl}_n\mathbb{R}, \mathfrak{o}(n), \mathfrak{so}(n)$ are well-defined, but only

$$exp : \mathfrak{so}(n) \to \mathrm{SO}(n) \tag{784}$$

is surjective.

Theorem 14.6

The Lie algebras for the following classical Lie groups are summarized as follows.

1. $\mathfrak{sl}_2\mathbb{C}$ is the real (or complex) vector space of traceless complex $n \times n$ matrices.
2. $\mathfrak{u}(n)$ is the real vector space of complex $n \times n$ skew-Hermitian matrices.
3. $\mathfrak{su}(n) = \mathfrak{u} \cap \mathfrak{sl}_2\mathbb{C}$. It is also a real vector space.
4. $\mathfrak{gl}_n\mathbb{C}$ is the real (or complex) vector space of complex $n \times n$ matrices.

Note that even though the matrices in these Lie algebras have complex coefficients, we have assigned them to be in a **real** vector space, which means that we are only allowed to take real linear combinations of these elements. That is, the field we are working over is $\mathbb{R}$ (this does not contradict any of the axioms for vector spaces). For example an element A in $\mathfrak{u}(n)$ or $\mathfrak{su}(n)$ must be anti-self adjoint, but iA is self adjoint.

Similarly, the Lie groups

$$\mathrm{GL}(n, \mathbb{C}), \mathrm{SL}(n, \mathbb{C}), \mathfrak{gl}_n\mathbb{C}, \mathfrak{sl}_n\mathbb{C} \tag{785}$$

are also smooth real manifolds embedded in $\mathrm{Mat}(n, \mathbb{C}) \simeq \mathbb{C}^{n^2} \simeq \mathbb{R}^{2n^2}$. So, we can view these four groups as manifolds embedded in $\mathbb{R}^{2n^2}$.

Note some of the similarities and differences between the real and complex counterparts of these Lie groups and algebras.

1. $\mathfrak{o}(n) = \mathfrak{so}(n)$, but $\mathfrak{u}(n) \neq \mathfrak{su}(n)$.
2. $exp : \mathfrak{gl}_n\mathbb{R} \to \mathrm{GL}(n, \mathbb{R})$ is not surjective, but $exp : \mathfrak{gl}_n\mathbb{C} \to \mathrm{GL}(n, \mathbb{C})$ is surjective due to the spectral theorem and surjectivity of $exp : \mathbb{C} \to \mathbb{C}^*$.
3. The exponential maps $exp : \mathfrak{u}(n) \to \mathrm{U}(n)$ and $exp : \mathfrak{su}(n) \to \mathrm{SU}(n)$ are surjective.

4. Still, $exp : \mathfrak{sl}_2\mathbb{C} \to \mathrm{SL}(2,\mathbb{C})$ is not surjective. This will be proved now.

Theorem 14.7

$exp : \mathfrak{sl}_2\mathbb{C} \to \mathrm{SL}(2,\mathbb{C})$ is not surjective.

Proof. Given $M \in \mathrm{SL}(n,\mathbb{C})$, assume that $M = e^A$ for some matrix $A \in \mathfrak{sl}_2\mathbb{C}$. Putting A into the Jordan Normal Form $J = NAN^{-1}$ means that J can either be of form

$$J = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} \lambda & 0 \\ 0 & -\lambda \end{pmatrix} \implies e^J = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} e^\lambda & 0 \\ 0 & e^{-\lambda} \end{pmatrix} \tag{786}$$

which is also in JNF in $\mathrm{SL}(2,\mathbb{C})$. But a matrix $P \in \mathrm{SL}(2,\mathbb{C})$ may exist with JNF of

$$K = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix} \tag{787}$$

which is not one of the 2 forms. So, $K \notin \mathrm{Im}\, exp \implies exp$ is not surjective.

Theorem 14.8

The exponential maps

$$exp : \mathfrak{u}(n) \to \mathrm{U}(n)$$
$$exp : \mathfrak{su}(n) \to \mathrm{SU}(n)$$

are surjective.

14.1.4 Lie Algebra of SE(n)

Recall that the group of affine rigid isometries is denoted $\mathrm{SE}(n)$. That is,

$$\mathrm{SE}(n) \equiv \mathrm{SO}(n) \ltimes \mathrm{Tran}\, \mathbb{R}^n \tag{788}$$

We can define the matrix representation of this affine transformation as such. Given an element $g \in \mathrm{SE}(n)$ such that

$$g(x) \equiv Rx + U,\ R \in \mathrm{SO}(n), U \in \mathrm{Tran}\, \mathbb{R}^n \tag{789}$$

we define the representation

$$\rho : \mathrm{SE}(n) \to \mathrm{GL}(n+1,\mathbb{R}), \rho(g) \equiv \begin{pmatrix} R & U \\ 0 & 1 \end{pmatrix} \tag{790}$$

where R is a real $n \times n$ matrix in $\mathrm{SO}(n)$ and U is a real n-vector in $\mathrm{Tran}\, \mathbb{R}^n \simeq \mathbb{R}^n$. We would then have

$$\rho(g) \begin{pmatrix} x \\ 1 \end{pmatrix} \equiv \begin{pmatrix} R & U \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix} = \begin{pmatrix} Rx + U \\ 1 \end{pmatrix} \in \mathbb{R}^{n+1} \tag{791}$$

Clearly, $\mathrm{SE}(n)$ is a Lie group, and the matrix representation ϱ of its Lie algebra $\mathfrak{se}(n)$ can be defined as the vector space of $(n+1) \times (n+1)$ matrices of the block form

$$A = \begin{pmatrix} \Omega & U \\ 0 & 0 \end{pmatrix} \tag{792}$$

where Ω is an $n \times n$ skew-symmetric matrix and $U \in \mathbb{R}^n$. Note that there are two different exponential maps here: one belonging to the abstract Lie group $\mathrm{SE}(n)$ and another belonging to the concrete, matrix group

$\mathrm{GL}(n+1,\mathbb{R})$. This can be represented with the commutative diagram.

$$\begin{array}{ccc} \mathfrak{se}(n) & \xrightarrow{exp} & SE(n) \\ \downarrow{\scriptstyle \varrho} & & \downarrow{\scriptstyle \rho} \\ \mathfrak{gl}_{n+1}\mathbb{R} & \xrightarrow{exp} & \mathrm{GL}(n+1,\mathbb{R}) \end{array}$$

Lemma 14.9

Given any $(n+1)\times(n+1)$ matrix of form

$$A = \begin{pmatrix} \Omega & U \\ 0 & 0 \end{pmatrix} \tag{793}$$

where Ω is any matrix and $U \in \mathbb{R}^n$,

$$A^k = \begin{pmatrix} \Omega^k & \Omega^{k-1}U \\ 0 & 0 \end{pmatrix} \tag{794}$$

where $\Omega^0 = I_n$, which implies that

$$e^A = \begin{pmatrix} e^\Omega & VU \\ 0 & 1 \end{pmatrix},\ V = I_n + \sum_{k\geq 1} \frac{\Omega^k}{(k+1)!} \tag{795}$$

Theorem 14.10

The exponential map

$$exp : \mathfrak{se}(n) \to SE(n) \tag{796}$$

is well-defined and surjective.

14.2 Representations of Lie Groups and Lie Algebras

Let $\mathcal{G}$ be an abstract group and let

$$\rho : \mathcal{G} \to \mathrm{GL}(V) \tag{797}$$

be the representation of $\mathcal{G}$. Then, let $\mathfrak{g}$ be the Lie algebra of $\mathcal{G}$, and $\mathfrak{gl}(V)$ be the Lie algebra of $\mathrm{GL}(V)$. Then, ρ induces another homomorphism

$$\varrho : \mathfrak{g} \to \mathfrak{gl}(V) \tag{798}$$

where the bracket structure (in this case, the comutator in the matrix algebra) is preserved.

$$\varrho([X,Y]) = [\varrho(X), \varrho(Y)] \tag{799}$$

We can visualize this induced homomorphism with the following commutative diagram, which states that $\rho \circ exp = exp \circ \varrho$.

$$\begin{array}{ccc} \mathcal{G} & \xrightarrow{\rho} & \mathrm{GL}(V) \\ {\scriptstyle exp}\uparrow & & \uparrow{\scriptstyle exp} \\ \mathfrak{g} & \xrightarrow{\varrho} & \mathfrak{gl}(V) \end{array}$$

Note that there are very crucial differences between ρ and ϱ. First, ρ is a homomorphism between **groups**, while ϱ is a homomorphism between **vector spaces**. Additionally, $\mathrm{GL}(V)$ is a group, not a linear space, while $\mathfrak{gl}(V)$ is a linear space. Finally, note that $\mathrm{GL}(V)$ is restricted to only matrices with nonzero determinants, while the elements of $\mathfrak{gl}(V)$ can be any matrix.

Example 14.5

The representation of $\mathrm{SE}(n)$ to $\mathrm{GL}(n+1\mathbb{R}$ and $\mathfrak{se}(n)$ to $\mathfrak{gl}_{n+1}\mathbb{R}$ induces the second homomorphism $\varrho : \mathfrak{gl}_{n+1}\mathbb{R} \to \mathrm{GL}(n+1, \mathbb{R})$.

Definition 14.11

The direct sum of representations is a representation. That is, if U is a representation and V is a representation, then $U \oplus V$ is a representation. That is, if

$$\rho_1 : \mathcal{G} \to U,\ \rho_1(g) = \begin{pmatrix} u_1 & u_2 \\ u_3 & u_4 \end{pmatrix} \tag{800}$$

and

$$\rho_2 : \mathcal{G} \to V,\ \rho_2(g) = \begin{pmatrix} v_1 & v_2 \\ v_3 & v_4 \end{pmatrix} \tag{801}$$

are two representations of the same group element $g \in \mathcal{G}$, then

$$(\rho_1 \oplus \rho_2) : \mathcal{G} \to (U \oplus V),\ (\rho_1 \oplus \rho_2)(g) = \begin{pmatrix} u_1 & u_2 & 0 & 0 \\ u_3 & u_4 & 0 & 0 \\ 0 & 0 & v_1 & v_2 \\ 0 & 0 & v_3 & v_4 \end{pmatrix} \tag{802}$$

is a bigger representation of g in $U \oplus V$.

Definition 14.12

V is irreducible if the only subspaces which are representations are only V and $\{0\}$.

For our case, we will consider that any representation can be written as a direct sum of irreducible representations. We will now proceed to find an irreducible representation of $\mathfrak{sl}_2\mathbb{C}$. This means that we want to find the smallest (lowest dimensional) vector space V such that there exists a representation

$$\varrho : \mathfrak{sl}_2\mathbb{C} \to \mathfrak{gl}(V) \tag{803}$$

We will write, as shorthand notation, that

$$H = \varrho(H), X = \varrho(X), Y = \varrho(Y) \tag{804}$$

Clearly, $H, X, Y \in \mathfrak{gl}(V) \simeq \mathfrak{gl}(\mathbb{C}^n)$. By the spectral theorem, we can find an orthonormal basis of eigenvectors $e_1, e_2, ..., e_n$ of the mapping H such that

$$He_i = \lambda_i e_i,\ \lambda_i \in \mathbb{C} \tag{805}$$

Since $[H, X] = 2X$, it follows that $HXe_i - XHe_i = 2Xe_i \implies H(Xe_i) = (\lambda_i + 2)(Xe_i) \implies Xe_i$ for all $i = 1, 2, ..., n$ are also eigenvectors of H with eigenvalue $(\lambda_i + 2)$, or $Xe_i = 0$. So, X is a "ladder operator" that maps each eigenvector e_i with eigenvalue λ_i to a different eigenvector e_j with eigenvalue $\lambda_j = \lambda_i + 2$. Having nowhere to be mapped to, the eigenvector with the largest eigenvalue (which must exist since V is finite dimensional) will get mapped to the 0 vector by X. Let us denote this eigenvector having the maximum eigenvalue m, as v_m.

Similarly, $[H, Y] = -2Y$ implies that

$$HYe_i - YHe_i = -2Ye_i \implies H(Ye_i) = (\lambda_i - 2)(Ye_i) \tag{806}$$

implying that Y maps each eigenvector e_i with eigenvaue λ_i to another eigenvector e_j with eigenvalue $\lambda_j = \lambda_i - 2$, except for the eigenvector with smallest eigenvalue, which gets mapped to 0. Since Y clearly

maps each eigenvector to a different eigenvector that has a strictly decreasing eigenvalue, we can construct a basis of V to be

$$\{v_m, Yv_m, Y^2v_m, Y^3v_m, ..., Y^{n-1}v_m\} \tag{807}$$

(remember that $Y^n v_m = 0$). So, elements of $\mathfrak{sl}_2\mathbb{C}$ acts on the space V with basis above. To continue, we introduce the following theorem.

Theorem 14.11

$$XY^j v_m = j(m-j+1)Y^{j-1}v_m \tag{808}$$

Proof. By induction on j using bracket relations.

V is n-dimensional. Since $Y^n v_m = 0$ and $Y^{n-1}v_m \neq 0$, we use the theorem above to get

$$0 = XY^n v_m = n(m-n+1)Y^{n-1}v_m \implies m-n+1 = 0 \tag{809}$$

So, $n = m+1$, which means that the eigenvalues of H are

$$m, m-2, m-4, \dots, m-2(n-1) = -m \tag{810}$$

and we are done. We now classify the 1, 2, and 3 dimensional irreducible representations of $\mathfrak{sl}_2\mathbb{C}$.

1. When $n = 1$ (i.e. dimension is 1), $m = n-1 = 0$, meaning that the greatest (and only) eigenvalue is 0. That is,

   $$Hv_0 = 0,\ Xv_0 = 0,\ Yv_0 = 0 \tag{811}$$

   which is the trivial representation of $\mathfrak{sl}_2\mathbb{C}$. Explicitly, we can completely define the representation (which is a linear homomorphism) with the three equations.

   $$\varrho(H) = (0),\ \varrho(X) = (0),\ \varrho(Y) = (0) \tag{812}$$

2. When $n = 2$ and $m = 1$. We now look for a 2 dimensional irreducible representation. The eigenvalues are 1 and -1, with $\{v_1, v_{-1}\}$ as a basis of 2 dimensional space V. Then we have

   $$\begin{aligned} Hv_1 = v_1,\ Hv_{-1} = -v_{-1} \\ Xv_1 = 0,\ Xv_{-1} = v_1 \\ Yv_1 = v_{-1},\ Yv_{-1} = 0 \end{aligned}$$

   which explicitly translates to the representation ϱ being defined

   $$\varrho(H) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \tag{813}$$

3. When $n = 3 \implies m = 2$, the basis is $\{v_{-2}, v_0, v_2\}$ with eigenvalues $2, 0, -2$, and the irreducible representation ϱ is defined

   $$\varrho(H) = \begin{pmatrix} 2 & & \\ & 0 & \\ & & -2 \end{pmatrix}, \varrho(Y) = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \varrho(X) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \tag{814}$$

4. The same process continues on for $n = 4, 5, ...$, and this entirely classifies the irreducible representations of $\mathfrak{sl}_2\mathbb{C}$.

14.2.1 Tensor Products of Group Representations

Definition 14.13

If V and W are two different representations of a group $\mathcal{G}$, then we know that $V \oplus W$ is also a representation of $\mathcal{G}$. Furthermore, the tensor product space $V \otimes W$ also defines a representation of $\mathcal{G}$. That is, given representations

$$\rho_V : \mathcal{G} \to \text{GL}(V)$$
$$\rho_W : \mathcal{G} \to \text{GL}(W)$$

The homomorphism $\rho_V \otimes \rho_W : \mathcal{G} \to \text{GL}(V \otimes W)$ is also a representation of $\mathcal{G}$, which is defined

$$(\rho_V \otimes \rho_W)(g)(v \otimes w) \equiv \rho_V(g)(v) \otimes \rho_W(g)(w) \tag{815}$$

or represented in shorthand notation,

$$g(v \otimes w) \equiv (gv) \otimes (gw) \tag{816}$$

We know that $\exp(H)$ acts on V and W since it is an element of $\text{GL}(V)$ and $\text{GL}(W)$. This means that

$$exp(H)(v \otimes w) \equiv \big(exp(H)(v)\big) \otimes \big(exp(H)(w)\big) \tag{817}$$

If H ($= \rho_V(H)$ or $\rho_W(H)$) has an eigenvalue λ on v in V and eigenvalue μ on w in W, then

$$exp(H)(v \otimes w) = (e^{\lambda} v) \otimes (e^{\mu} w) = e^{\lambda + \mu} v \otimes w \tag{818}$$

That is, eigenvalues of H **add** on tensor products.

Example 14.6

Recall that the 2 dimensional representation V of $\mathfrak{sl}_2\mathbb{C}$ has eigenvalues 1 and -1 (with corresponding eigenvectors e_1 and e_{-1}). So, $V \otimes V$ has eigenvalues

$$(-1) + (-1) = -2, \;\; (-1) + 1 = 0$$
$$1 + (-1) = 0, \;\; 1 + 1 = 2$$

Therefore, the eigenvalues of $V \otimes V$ is -2 (geometric multiplicity of 1), 0 (geometric multiplicity of 2), and 2 (geometric multiplicity of 1), (Notation-wise, the n-dimensional irreducible representation of $\mathfrak{sl}_2\mathbb{C}$ is denoted $\mathbf{n}$.) which means that

$$\mathbf{2} \otimes \mathbf{2} = \mathbf{3} \oplus \mathbf{1} \tag{819}$$

We can decompose $V \otimes V$ into its symmetric and exterior power components. $\text{Sym}^2 V$ has basis (of eigenvectors)

$$\{e_{-1} \odot e_{-1},\; e_{-1} \odot e_1,\; e_1 \odot e_1\} \tag{820}$$

where the corresponding eigenvalues are -2, 0, and 2, respectively. So, $\dim Sym^2 V = 3$, which means that $Sym^2 V = \mathbf{3}$. As for the exterior power component of V, $\Lambda^2 V$ has basis $\{e_{-1} \wedge e_1\}$ with eigenvalue $= 0 \implies \dim \Lambda^2 V = 1$, meaning that $\Lambda^2 V = \mathbf{1}$. Therefore,

$$V \otimes V = Sym^2 V \oplus \Lambda^2 V = \mathbf{3} \oplus \mathbf{1} \tag{821}$$

14.3 Topological Decompositions of Lie Groups

Definition 14.14

Let us define
1. $\mathrm{S}(n)$ is the vector space of real, symmetric $n \times n$ matrices.
2. $\mathrm{SP}(n)$ is the set of symmetric, positive semidefinite matrices.
3. $\mathrm{SPD}(n)$ is the set of symmetric, positive definite matrices.

Note that $\mathrm{SP}(n)$ and $\mathrm{SPD}(n)$ are not even vector spaces at all.

Lemma 14.12

The exponential map

$$exp : S(n) \to SPD(n) \tag{822}$$

is a homeomorphism. One may be tempted to call $\mathrm{S}(n)$ the Lie algebra of $\mathrm{SPD}(n)$, but this is not the case. $\mathrm{S}(n)$ is not even a Lie algebra since the commutator is not algebraically closed. Furthermore, $\mathrm{SPD}(n)$ is not even a multiplicative group (since matrix multiplication is not closed).

Recall from linear algebra the Polar Decomposition. We express this result in a slightly modified way.

Theorem 14.13 (Polar Decomposition)

Given a Euclidean space $\mathbb{E}^n$ and any linear endomorphism f of $\mathbb{E}^n$, there are two positive definite self-adjoint linear maps $h_1, h_2 \in \mathrm{End}(\mathbb{E}^n)$ and $g \in \mathrm{O}(n)$ such that

$$f = g \circ h_1 = h_2 \circ g \tag{823}$$

That is, such that f can be decomposed into the following compositions of functions that commute.

$$\begin{array}{ccc} \mathbb{E}^n & \xrightarrow{h_2} & \mathbb{E}^n \\ {\scriptstyle g}\uparrow & \nearrow_{f} & \uparrow{\scriptstyle g} \\ \mathbb{E}^n & \xrightarrow{h_1} & \mathbb{E}^n \end{array}$$

This means that there is a bijection between $Mat(n, \mathbb{R})$ and $O(n) \times SP(n)$. If f is an automorphism, then this decomposition is unique.

Corollary 14.14

The two topological groups are homeomorphic.

$$GL(n, \mathbb{R}) \cong O(n) \times SPD(n) \tag{824}$$

Corollary 14.15

For every invertible real matrix $A \in GL(n, \mathbb{R})$, there exists a unique orthogonal matrix R and unique symmetric matrix S such that

$$A = Re^S \tag{825}$$

$\implies$ there is a bijection between $\mathrm{GL}(n, \mathbb{R})$ and $O(n) \times S(n) \simeq \mathbb{R}^{n(n+1)/2}$. Moreover, they are homeomorphic. That is,

$$\mathrm{GL}(n, \mathbb{R}) \simeq O(n) \times S(n) \simeq O(n) \times \mathbb{R}^{n(n+1)/2} \tag{826}$$

This essentially reduces the study of $\mathrm{GL}(n, \mathbb{R})$ to the study of $O(n)$, which is nice since $O(n)$ is compact.

Corollary 14.16

Given a real matrix A, if $\det A > 0$, then we can decompose A as

$$A = Re^S \tag{827}$$

where $R \in SO(n)$ and $S \in S(n)$.

Corollary 14.17

There exists a bijection between

$$\mathrm{SL}(n, \mathbb{R}) \text{ and } SO(n) \times (S(n) \cap \mathfrak{sl}_n \mathbb{R}) \tag{828}$$

Proof. $A \in \mathrm{SL}(n, \mathbb{R}) \implies 1 = \det A = \det R \det e^S = \det e^S \implies \det e^S = e^{\mathrm{Tr}\, S} = 1 \implies \mathrm{Tr}\, S = 0 \implies S \in \mathrm{S}(n) \cap \mathfrak{sl}_n \mathbb{R}$.

Definition 14.15

Let us define
1. $\mathrm{H}(n)$ is the real vector space of $n \times n$ Hermitian matrices.
2. $\mathrm{HP}(n)$ is the set of Hermitian, positive semidefinite $n \times n$ matrices.
3. $\mathrm{HPD}(n)$ is the set of Hermitian, positive definite $n \times n$ matrices.

Similarly, $\mathrm{HP}(n)$ and $\mathrm{HPD}(n)$ are not vector space. They are just sets.

Lemma 14.18

The exponential mapping

$$exp : H(n) \to HPD(n) \tag{829}$$

is a homeomorphism.

However again, $\mathrm{HPD}(n)$ is not a Lie group (multiplication is not algebraically closed) nor is $\mathrm{H}(n)$ a Lie algebra (commutator is not algebraically closed). By the polar form theorem of complex $n \times n$ matrices, we have a (not necessarily unique) bijection between

$$\mathrm{Mat}(n, \mathbb{C}) \text{ and } U(n) \times HP(n) \tag{830}$$

which implies that

$$\mathrm{GL}(n, \mathbb{C}) \cong U(n) \times HPD(n) \tag{831}$$

Corollary 14.19

For every complex invertible matrix A, there exists a unique decomposition

$$A = Ue^S \tag{832}$$

where $U \in U(n)$ and $S \in H(n)$, which implies that the following groups are homeomorphic.

$$\begin{aligned} \mathrm{GL}(n, \mathbb{C}) &\cong U(n) \times H(n) \\ &\cong U(n) \times \mathbb{R}^{n^2} \end{aligned}$$

This essentially reduces the study of $\mathrm{GL}(n, \mathbb{C})$ to that of $\mathrm{U}(n)$.

Corollary 14.20

There exists a bijection between

$$\mathrm{SL}(n, \mathbb{C}) \text{ and } SU(n) \times (H(n) \cap \mathfrak{sl}_n \mathbb{C}) \tag{833}$$

Proof. Similarly, when $A = Ue^S$, we know that $|\det U| = 1$ and $\operatorname{Tr} S$ is real (since by the Spectral theorem, every self adjoint matrix has a real spectral decomposition). Since S is Hermitian, this implies that $\det e^S > 0$. If $A \in \mathrm{SL}(n, \mathbb{C})$, then $\det A = 1 \implies \det e^S = 1 \implies S \in H(n) \cap \mathfrak{sl}_n \mathbb{C}$.

14.4 Linear Lie Groups

We will assume that the reader has the necessary background knowledge in manifolds, chart mappings, diffeomorphisms, tangent spaces, and transition mappings.

Recall that the algebra of real $n \times n$ matrices $\mathrm{Mat}(n, \mathbb{R})$ is bijective to $\mathbb{R}^{n^2}$, which is a topological space. Therefore, this bijection

$$i : (\mathbb{R}^{n^2}, \tau_E) \to \mathrm{Mat}(n, \mathbb{R}) \tag{834}$$

induces a topology on $\mathrm{Mat}(n, \mathbb{R})$, defined

$$\tau_M \equiv \{U \in \mathrm{Mat}(n, \mathbb{R}) \mid e^{-1}(U) \in \tau_E\} \tag{835}$$

With this, consider the subset

$$\mathrm{GL}(n, \mathbb{R}) \subset \mathrm{Mat}(n, \mathbb{R}) \tag{836}$$

where

$$\mathrm{GL}(n, \mathbb{R}) \equiv \{x \in \mathrm{Mat}(n, \mathbb{R}) \mid \det x \neq 0\} \tag{837}$$

This set, as we expect, is a multiplicative group.

Definition 14.16

The **general linear group**, denoted $\mathrm{GL}(n, \mathbb{R})$ is the set of $n \times n$ matrices with nonzero determinant. The more technical definition is that $\mathrm{GL}(n, \mathbb{R})$ is really just the automorphism group of $\mathbb{R}^n$,

$$\mathrm{GL}(n, \mathbb{R}) \equiv \mathrm{Aut}(\mathbb{R}^n) \tag{838}$$

but it is customary to assume a basis on $\mathbb{R}^n$ in order to realize $\mathrm{GL}(n, \mathbb{R})$ as a matrix group. Note that the procedure of assuming a basis on $\mathbb{R}^n$ is the same as defining a representation of the abstract group $\mathrm{GL}(n, \mathbb{R})$. Both assigns a real $n \times n$ matrix to each element of $\mathrm{GL}(n, \mathbb{R})$.

In this way, we can view $\mathrm{GL}(n, \mathbb{R})$ as a topological space in $\mathbb{R}^{n^2}$, and it is fine to interpret $\mathrm{GL}(n, \mathbb{R})$ as a matrix group rather than an abstract group.

Since the matrix representation of $\mathrm{GL}(n, \mathbb{R})$ is always well defined, the abstract subgroups of $\mathrm{GL}(n, \mathbb{R})$, which are $\mathrm{SL}(n, \mathbb{R}), O(n)$, and $SO(n)$, also have well defined matrix representations (that we are all familiar with). Additionally, since there exists a bijection

$$\mathrm{Mat}(n, \mathbb{C}) \cong \mathbb{C}^{n^2} \cong \mathbb{R}^{2n^2} \tag{839}$$

we can view $\mathrm{GL}(n, \mathbb{C})$ as a subset of $\mathbb{R}^{2n^2}$, meaning that the subgroups $\mathrm{SL}(n, \mathbb{C}), U(n)$, and $SU(n)$ of $\mathrm{GL}(n, \mathbb{C})$ can also be viewed as subsets of $\mathbb{R}^{2n^2}$. This also applies to $SE(n)$ since it is a subgroup of $\mathrm{SL}(n+1, \mathbb{R})$. We formally state it now.

Theorem 14.21

SE(n) is a linear Lie group.

Proof. The matrix representation of elements $g \in SE(n)$ is

$$\rho(g) \equiv \begin{pmatrix} R_g & U_g \\ 0 & 1 \end{pmatrix},\ R_g \in SO(n), U_g \in \mathbb{R}^n \tag{840}$$

But such matrices also belong to the bigger group $\text{SL}(n+1,\mathbb{R}) \implies SE(n) \subset \text{SL}(n+1,\mathbb{R})$. Moreover, this canonical embedding

$$i : SE(n) \to \text{SL}(n+1,\mathbb{R}) \tag{841}$$

is a group homomorphism since

$$\begin{aligned} i\big(\rho(g_1 \cdot g_2)\big) &= \begin{pmatrix} RS & RV+U \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} R & U \\ 0 & 1 \end{pmatrix}\begin{pmatrix} S & V \\ 0 & 1 \end{pmatrix} = \rho\big(i(g_1)\cdot i(g_2)\big) \end{aligned}$$

and the inverse is given by

$$\begin{pmatrix} R & U \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} R^{-1} & -R^{-1}U \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} R^T & -R^TU \\ 0 & 1 \end{pmatrix} \tag{842}$$

is also consistent between the inverse operation in SE(n) and SL$(n+1,\mathbb{R})$. Therefore, SE(n) is a subgroup of SL$(n+1,\mathbb{R})$, which is a subgroup of GL$(n+1,\mathbb{R})$.

Note that even though SE(n) is diffeomorphic (a topological relation) to SO$(n) \times \mathbb{R}^n$, it is **not** isomorphic (an algebraic relation) sicne group operations are not preserved. Therefore, we write this "equality" as a semidirect product of groups.

$$SE(n) \equiv SO(n) \ltimes \mathbb{R}^n \tag{843}$$

Therefore, all of the classical Lie groups that we have mentioned can be viewed as subsets of $\mathbb{R}^N$ (with the subspace topology) and as subgroups of GL$(N,\mathbb{R})$ for some big enough N. This defines a special family of Lie groups, called linear Lie groups.

Definition 14.17

A **linear Lie group** is a subgroup of GL$(n,\mathbb{R})$ for some $n \geq 1$ which is also a smooth manifold in $\mathbb{R}^{n^2}$.

Theorem 14.22 (Von Neumann, Cartan)

A closed subgroup $\mathcal{G}$ of GL$(n,\mathbb{R})$ is a linear Lie group. That is, a closed subgroup $\mathcal{G}$ of GL$(n,\mathbb{R})$ is a smooth manifold in $\mathbb{R}^{n^2}$.

Definition 14.18

Since a linear Lie group $\mathcal{G}$ is a smooth submanifold in $\mathbb{R}^N$, we can take its tangent space at the identity element I, which is defined

$$T_I\mathcal{G} \equiv \{p'(0) \mid p : I \subset \mathbb{R} \to \mathcal{G}, p(0) = I\} \tag{844}$$

where p is a path function on $\mathcal{G}$.

Note that we haven't mentioned anything about the exponential map up to now. We mention the relationship between this map and the Lie algebra with the following theorem.

Theorem 14.23

Let $\mathcal{G}$ be a linear Lie group. The set $\mathfrak{g}$ defined such that

$$\mathfrak{g} \equiv \{X \in \text{Mat}(n, \mathbb{R}) \mid e^{tX} \in \mathcal{G} \; \forall t \in \mathbb{R}\} \tag{845}$$

is equal to the tangent space of $\mathcal{G}$ at the identity element. That is,

$$\mathfrak{g} = T_I \mathcal{G} \tag{846}$$

Furthermore, $\mathfrak{g}$ is closed under the commutator

$$[A, B] \equiv AB - BA \tag{847}$$

This theorem ensures that given a linear Lie group $\mathcal{G}$, the tangent space $\mathfrak{g}$ exists and is closed under the commutator. We formally define this space.

Definition 14.19

The Lie algebra of a linear Lie group is a real vector space (of matrices) together with a algebraically closed bilinear map

$$[A, B] \equiv AB - BA \tag{848}$$

called the **commutator**.

The definition of $\mathfrak{g}$ given in the previous theorem shows that

$$exp : \mathfrak{g} \to \mathcal{G} \tag{849}$$

is well defined. In general, exp is neither injective nor surjective. Visually, this exponential mapping is what connects the Lie algebra, i.e. the tangent space of manifold $\mathcal{G}$ to the actual Lie group $\mathcal{G}$. To define the inverse map that maps Lie group elements to Lie algebra ones, we can simply just compute the tangent vectors of the manifold $\mathcal{G}$ at the identity I by taking the derivative of arbitrary path functions in $\mathcal{G}$. That is, for every $X \in T_I \mathcal{G}$, we define the smooth curve

$$\gamma_X : t \mapsto e^{tX} \tag{850}$$

where $\gamma_X(0) = I$. If we take the derivative of this curve, with respect to t at $t = 0$, we will get the tangent vector X corresponding to that group element $g = e^X$. More visually, we just need to take the collection of all smooth path functions γ on manifold $\mathcal{G}$ such that $\gamma(0) = I$. Then, taking the derivative of all these paths at $t = 0$ will produce the collection of all tangent vectors at the identity element. We show this process in the following examples.

Theorem 14.24

The matrix representation of $\mathfrak{sl}_n\mathbb{R}$ is precisely the set of traceless $n \times n$ matrices.

Proof. Clearly, $\mathfrak{sl}_n\mathbb{R}$ is a vector space since it is a Lie algebra. So, $X \in \mathfrak{sl}_n\mathbb{R} \implies tX \in \mathfrak{sl}_n\mathbb{R}$ for all

$t \in \mathbb{R} \implies \det e^{tX} = 1$ for all $t \in \mathbb{R}$, for all $X \in \mathfrak{sl}_n\mathbb{R}$. But we use the identity

$$\begin{aligned} \det e^{tX} = e^{\operatorname{Tr}(tX)} &\implies 1 = e^{\operatorname{Tr}(tX)} \\ &\implies \operatorname{Tr}(tX) = 0 \\ &\implies \operatorname{Tr}(X)t = 0 \implies \operatorname{Tr} X = 0 \end{aligned}$$

We now provide an alternative, better proof. We first need a lemma.

Lemma 14.25

$\det'(I) = \operatorname{Tr}$. That is, the differential of the det operator, evaluated at the identity matrix, is equal to the trace. That is, given any matrix T in the vector space of matrices,

Proof.

$$\begin{aligned} \det'(I)(T) = \nabla_T \det(I) & \\ &= \lim_{\varepsilon \to 0} \frac{\det(I + \varepsilon T) - \det I}{\varepsilon} \\ &= \lim_{\varepsilon \to 0} \frac{\det(I + \varepsilon T) - 1}{\varepsilon} \end{aligned}$$

Clearly, $\det(I + \varepsilon(T)) \in \mathbb{R}[\varepsilon]$, where the constant term of the polynomial approaches 1 and the linear term (coefficient of ε) is $\operatorname{Tr} T$. So,

$$\nabla_T \det I = \lim_{\varepsilon \to 0} ... + \operatorname{Tr} T = \operatorname{Tr} T \tag{851}$$

This means that the instantaneous rate at which det changes at I when traveling in direction T is directly

Proof. Let $R : \mathbb{R} \to \mathrm{SL}(n, \mathbb{R})$ such that $R(0) = I$. Then, by definition, $\operatorname{Im} R \subset \mathrm{SL}(n, \mathbb{R}) \implies \det(R(t)) = 1$ for all $t \in (-\varepsilon, \varepsilon)$. Compute the derivative of the mapping $\det \circ R$.

$$\begin{aligned} (\det \circ R)(t) = 1 &\implies \det'\big(R(t)\big) \cdot R'(t) \\ &\implies \det'(I) = \det'\big(R(t)\big) = 0 \end{aligned}$$

We now use the previous lemma get that

$$\det'\big(R'(0)\big) = \det'(I) = 0 \implies \operatorname{Tr} R'(0) = 0 \tag{852}$$

Theorem 14.26

The matrix representation of $\mathfrak{so}(n)$ is precisely the set of antisymmetric matrices.

Proof. Let $R : \mathbb{R} \to SO(n)$ be a arbitrary smooth curve in $\mathrm{SL}(n)$ such that $R(0) = I$. Then, for all $t \in (-\epsilon, \epsilon)$,

$$R(t)R(t)^T = I \tag{853}$$

Taking the derivative at $t = 0$, we get

$$R'(0)R(0)^T + R(0)R'(0)^T = 0 \implies R'(0) + R'(0)^T = 0 \tag{854}$$

which states that the tangent vector $X = R'(0)$ is skew symmetric. Since the diagonal elements of a skew symmetric matrix are 0, the trace is 0 and the condition that $\det R = 1$ yields nothing new. This shows that $\mathfrak{o}(n) = \mathfrak{so}(n)$.

We have only worked with linear Lie groups so far. The reason that linear Lie groups are so nice to work with is because they have well defined matrix representations. This allows us to have concrete structures on these groups and their Lie algebras.

1. A linear Lie group is concretely defined as a submanifold of $\mathbb{R}^N$, while a general one is an abstract manifold.
2. The Lie bracket with regards to a linear Lie group is defined to be the commutator
$$[A, B] \equiv AB - BA \tag{855}$$
but for elements that are not matrices this doesn't make sense.
3. The exponential map from the algebra to the group is defined
$$e^A \equiv \sum_{k=0}^{\infty} \frac{1}{k!} A^k \tag{856}$$
but if A is not a matrix, then exp cannot be defined this way.

We seek to generalize these concepts to abstract Lie groups, but we will do this in the next section.

14.4.1 Lie Algebras of SO(3) and SU(2), Revisited

Example 14.7

The Lie algebra $\mathfrak{so}(3)$ is the real vector space of 3×3 skew symmetric matrices of form
$$\begin{pmatrix} 0 & -d & c \\ d & 0 & -b \\ -c & b & 0 \end{pmatrix} \tag{857}$$
where $b, c, d \in \mathbb{R}$. The Lie bracket $[A, B]$ of $\mathfrak{so}(3)$ is also just the usual commutator.
We can define an isomorphism of Lie algebras $\psi : (\mathbb{R}^3, \times) \to \mathfrak{so}(3)$ (where $\times$ is the cross product) by the formula
$$\psi(b, c, d) \equiv \begin{pmatrix} 0 & -d & c \\ d & 0 & -b \\ -c & b & 0 \end{pmatrix} \tag{858}$$
where, by definition,
$$\psi(u \times v) = [\psi(u), \psi(v)] \tag{859}$$
It is also easily verified that for all $u, v \in \mathbb{R}^3$,
$$\psi(u)(v) = u \times v \tag{860}$$

Example 14.8

Similarly, we can see that $\mathfrak{su}(2)$ is the real vector space consisting of all complex 2×2 skew Hermitian matrices of null trace, which is of form
$$i(d\sigma_1 + c\sigma_2 + b\sigma_3) = \begin{pmatrix} ib & c + id \\ -c + id & -ib \end{pmatrix} \tag{861}$$

where $\sigma_1, \sigma_2, \sigma_3$ are the Pauli spin matrices. We can also define an isomorphism of Lie algebras $\varphi : (\mathbb{R}^3, \times) \to \mathfrak{su}(2)$ by the formula

$$\varphi(b,c,d) = \frac{i}{2}(d\sigma_1 + c\sigma_2 + b\sigma_3) = \frac{1}{2}\begin{pmatrix} ib & c+id \\ -c+id & -ib \end{pmatrix} \tag{862}$$

where, by definition of ismorphism, we have

$$\varphi(u \times v) = [\varphi(u), \varphi(v)] \tag{863}$$

We now restate the connection between the groups SO(3) and SU(2). Note that letting $\theta = \sqrt{b^2+c^2+d^2}$, we can write

$$A = \frac{1}{\theta}(d\sigma_1 + c\sigma_2 + b\sigma_3) = \frac{1}{\theta}\begin{pmatrix} ib & c+id \\ -c+id & -ib \end{pmatrix} \tag{864}$$

such that $A^2 = I$. With this, we can rewrite the exponential map as

$$exp : \mathfrak{su}(2) \to SU(2),\ exp(i\theta A) = \cos\theta I + i\sin\theta A \tag{865}$$

As for the isomorphism $\varphi : (\mathbb{R}^3, \times) \to \mathfrak{su}(2)$, we have

$$\varphi(b,c,d) \equiv \frac{1}{2}\begin{pmatrix} ib & c+id \\ -c+id & -ib \end{pmatrix} = i\frac{\theta}{2}A \tag{866}$$

Similarly, we can view the exponential map $exp : (\mathbb{R}^3, \times) \to SU(2)$ as

$$exp(\theta v) = \tag{867}$$

Example 14.9

The lie algebra $\mathfrak{se}(n)$ is the set of all matrices of form

$$\begin{pmatrix} B & U \\ 0 & 0 \end{pmatrix} \tag{868}$$

where $B \in \mathfrak{so}(n)$ and $U \in \mathbb{R}^n$. The Lie bracket is given by

$$\begin{pmatrix} B & U \\ 0 & 0 \end{pmatrix}\begin{pmatrix} C & V \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} C & V \\ 0 & 0 \end{pmatrix}\begin{pmatrix} B & U \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} BC-CB & BV-CU \\ 0 & 0 \end{pmatrix} \tag{869}$$

14.5 Abstract Lie Groups

Definition 14.20

A (real) **Lie group** $\mathcal{G}$ is a group $\mathcal{G}$ that is also a real, finite-dimensional smooth manifold where group multiplication and inversion are smooth maps.

Definition 14.21

A (real) Lie algebra $\mathfrak{g}$ is a real vector space with a map

$$[\cdot,\cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g} \tag{870}$$

called the Lie bracket satisfying bilinearity, antisymmetricity, and the Jacobi Identity.

To every Lie group $\mathcal{G}$ we can associate a Lie algebra $\mathfrak{g}$ whose underlying vector space is the tangent space of $\mathcal{G}$ at the identity element. Additionally, the exponential map allows us to map elements from the Lie algebra to the Lie group. These concrete definitions in the context of linear Lie groups is easy to work with, but has some minor problems: to use it we first need to represent a Lie group as a group of matrices, but not all Lie groups can be represented in this way.

To do this, we must introduce further definitions.

Definition 14.22

Let M_1 (m_1-dimensional) and M_2 (m_2 dimensional) be manifolds in $\mathbb{R}^N$. For any smooth function $f : M_1 \to M_2$ and any $p \in M_1$, the function

$$f_p' : T_pM_1 \to T_{f(p)}M_2 \tag{871}$$

called the **tangent map, derivative, or differential** of f at p, is defined as follows. For every $v \in T_pM_1$ and every smooth curve $\gamma : I \to M_1$ such that $\gamma(0) = p$ and $\gamma'(0) = v$,

$$f_p'(v) \equiv (f \circ \gamma)'(0) \tag{872}$$

The map f_p' is also denoted df_p and is a linear map.

Definition 14.23

Given two Lie groups $\mathcal{G}_1$ and $\mathcal{G}_2$, a **homomorphism of Lie groups** is a function

$$f : \mathcal{G}_1 \to \mathcal{G}_2 \tag{873}$$

that is both a group homomorphism and a smooth map (between manifolds $\mathcal{G}_1$ and $\mathcal{G}_2$). An **isomorphism of Lie groups** is a bijective function f such that both f and f^{-1} are homomorphisms of Lie groups.

Definition 14.24

Given two Lie algebras $\mathfrak{g}_1$ and $\mathfrak{g}_2$, a **homomorphism of Lie algebras** is a function

$$f : \mathfrak{g}_1 \to \mathfrak{g}_2 \tag{874}$$

that is a linear homomorphism that preserves Lie brackets; that is,

$$f([A, B]) = [f(A), f(B)] \tag{875}$$

for all $A, B \in \mathfrak{g}$. An **isomorphism of Lie algebras** is a bijective function $f$such that both f and f^{-1} are homomorphisms of Lie algebras.

Theorem 14.27

If $f : \mathcal{G}_1 \to \mathcal{G}_2$ is a homomorphism of Lie groups, then

$$f_I' : \mathfrak{g}_1 \to \mathfrak{g}_2 \tag{876}$$

is a homomorphism of Lie algebras.

We have explained how to construct the Lie bracket (as the commutator) of the Lie algebra of a linear Lie

group, but we have not defined how to construct the Lie bracket for general Lie groups. There are several ways to do this, and we describe one such way through **adjoint representations**.

Definition 14.25

Given a Lie group $\mathcal{G}$, we define a **left translation** as the map

$$L_a : \mathcal{G} \to \mathcal{G},\ L_a(b) \equiv ab \tag{877}$$

for all $b \in \mathcal{G}$. Similarly, the **right translation** is defined

$$R_a : \mathcal{G} \to \mathcal{G},\ R_a(b) \equiv ba \tag{878}$$

for all $b \in \mathcal{G}$.

Both L_a and R_a are diffeomorphisms. Additionally, given the automorphism

$$R_{a^{-1}} L_a \equiv R_{a^{-1}} \circ L_a,\ R_{a^{-1}} L_a(b) \equiv aba^{-1} \tag{879}$$

the derivative

$$(R_{a^{-1}} L_a)'_I : \mathfrak{g} \to \mathfrak{g} \tag{880}$$

is an ismorphism of Lie algebras, also denoted

$$\text{Ad}_a : \mathfrak{g} \to \mathfrak{g} \tag{881}$$

Definition 14.26

This induces another map $a \mapsto \text{Ad}_a$, which is a map of Lie groups

$$Ad : \mathcal{G} \to \text{GL}(\mathfrak{g}) \tag{882}$$

which is called the **adjoint representation of** $\mathcal{G}$. In the case of a linear map, we can verify that

$$\text{Ad}(a)(X) \equiv \text{Ad}_a(X) \equiv aXa^{-1} \tag{883}$$

for all $a \in \mathcal{G}$ and for all $X \in \mathfrak{g}$.

Definition 14.27

Furthermore, the derivative of this map at the identity

$$\text{Ad}'_I : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g}) \tag{884}$$

is a map between Lie algebras, denoted simply as

$$\text{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g}) \tag{885}$$

called the **adjoint representation** of $\mathfrak{g}$. It is easily visualized with the following commutative diagram.

$$\begin{array}{ccc} \mathcal{G} & \xrightarrow{Ad} & \text{GL}(\mathfrak{g}) \\ {\scriptstyle exp}\uparrow & & \uparrow{\scriptstyle exp} \\ \mathfrak{g} & \xrightarrow{ad} & \mathfrak{gl}(\mathfrak{g}) \end{array}$$

We define the map ad to be

$$ad(A)(B) \equiv [A, B] \tag{886}$$

where $[A, B]$ is the Lie bracket (of $\mathfrak{g}$) of $A, B \in \mathfrak{g}$. We can actually conclude something stronger about this mapping. Since the Lie bracket of $\mathfrak{g}$ satisfies the properties of the bracket, the Jacobi identity of $[\cdot, \cdot]$ implies that ad is a Lie algebra homomorphism.

$$\begin{aligned}
&[x,[y,z]] + [y,[z,x]] + [z,[x,y]] = 0 \\
\implies &[x, ad(y)(z)] + [y, ad(z)(x)] + [z, ad(x)(y)] = 0 \\
\implies &ad(x)\big(ad(y)(z)\big) + ad(y)\big(ad(z)(x)\big) + ad(z)\big(ad(x)(y)\big) = 0 \\
\implies &ad(x)ad(y)(z) - ad(y)ad(x)z - ad\big(ad(x)(y)\big)(z) = 0 \\
\implies &\big(ad(x)ad(y) - ad(y)ad(x)\big)(z) = ad\big(ad(x)(y)\big)(z) \\
\implies &[ad(x), ad(y)](z) = ad([x,y])(z) \\
\implies &[ad(x), ad(y)] = ad([x,y])
\end{aligned}$$

Therefore, ad preserves brackets and thus ad is a Lie algebra homomorphism. That is,

$$\text{ad}([A,B]) = [\text{ad}(A), \text{ad}(B)] \tag{887}$$

Note that the bracket on the left side represents the bracket of $\mathfrak{g}$, while the bracket on the right represents the Lie bracket from the Lie algebra $\mathfrak{gl}(\mathfrak{g})$. The fact that ad is a Lie algebra homomorphism indicates that it is a representation of $\mathfrak{g}$, which is why it's called the adjoint representation.

Definition 14.28

This construction finally allows us to define the Lie bracket in the case of a general Lie group. The Lie bracket on $\mathfrak{g}$ is defined as

$$[A, B] \equiv \text{ad}(A)(B) \tag{888}$$

We would also need to introduce a general exponential map for non-linear Lie groups, but we will not do it here.

14.6 Exercises