



ABSTRACT ALGEBRA

MATH1530

PROFESSOR JORDAN KOSTIUK

Brown University



EDITED BY

RICHARD TANG



Contents

1	Set Theory	3
1.1	Sets	3
1.1.1	The Well-Ordering Principle	4
1.2	Functions	5
2	Groups: Part I	7
2.1	Motivation	7
2.1.1	Permutations	7
2.2	(Abstract) Groups	8
2.2.1	Examples of Groups	10
2.2.2	Cyclic Groups	11
2.3	Group Homomorphisms	11
2.4	Subgroups, Cosets, and Lagrange's Theorem	14
2.4.1	Cosets	16
2.5	Products of Groups	19
3	Rings: Part I	21
3.1	Review of Number Theory	21
3.1.1	Equivalence Relations	21
3.1.2	Modular Arithmetic	22
3.2	Abstract Rings and Ring Homomorphisms	23
3.3	Interesting Examples of Rings	25
3.4	Important Properties of Rings	27
3.5	Unit Groups and Product Rings	28
3.5.1	Unit Groups	28
3.5.2	Product Rings	30
3.6	Ideals and Quotient Rings	31
3.7	Prime Ideals and Maximal Ideals	37
4	Vector Spaces: Part 1	40
4.1	Introduction to Vector Spaces	40
4.2	Vector Spaces and Linear Transformations	40
4.3	Interesting Examples of Vector Spaces	42
4.4	Bases and Dimension	43
5	Fields: Part I	49
5.1	Introduction to Fields	49
5.2	Abstract Fields and Homomorphisms	49
5.3	Interesting Examples of Fields	50
5.4	Subfields and Extension Fields	51
5.5	Polynomial Rings	54

<i>CONTENTS</i>	2
5.6 Building Extension Fields	56
5.7 Finite Fields	60

Chapter 1

Set Theory

Set theory forms a basis for all of higher mathematics. We begin with a brief introduction.

§1.1 Sets

Definition 1.1.1: Sets

A **set** is a (possibly empty) collection of elements. If S is a set and a is some object, then a is either an element of S or not. We write:

- $a \in S$ if a is an element of S .
- $a \notin S$ if a is not an element of S .

The empty set is denoted \emptyset . We use $|S|$ or $\#S$ to denote the cardinality (number of elements) in a finite set.

Definition 1.1.2: Natural Numbers

The **natural numbers** are the set

$$\mathbb{N} = \{1, 2, \dots\}.$$

Formally, we define \mathbb{N} as follows:

1. \mathbb{N} contains an initial element 1.
2. $\forall n \in \mathbb{N}$, there is an incremental rule that creates the next element $n + 1$.
3. We can reach every element of \mathbb{N} by starting with 1 and repeatedly adding 1.

Remark 1. \mathbb{N} is totally ordered. We say m is less than n if n appears before m when we start from 1 and add repeatedly. In this case we write $m < n$ or $m \leq n$ if $m = n$.

Example 1. Let

$$\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$$

denote the set of integers, and

$$Q = \left\{ \frac{a}{b} \mid a, b \in \mathbb{Z}, b \neq 0 \right\}.$$

the set of rationals.

Definition 1.1.3: Set Operations

Let S, T be sets.

1. S is a **subset** of T if every element of S is an element of T , i.e. $a \in S \rightarrow a \in T$. We write

$$S \subset T.$$

2. The **union** of S and T is the set of elements that belong to S or belong to T , denoted

$$S \cup T = \{a \mid a \in S \text{ or } a \in T\}.$$

3. The **intersection** of S and T is the set of elements that belong to both S and T , denoted

$$S \cap T = \{a \mid a \in S \text{ and } a \in T\}.$$

4. If $S \subset T$, the **complement** of S in T is the set of elements in T not in S :

$$S^c = T - S = T \setminus S = \{a \in T \mid a \notin S\}.$$

5. The **product** of S and T is the set of ordered pairs

$$S \times T = \{(a, b) \mid a \in S, b \in T\}.$$

We have projection maps

$$\begin{aligned} \text{proj}_1 : S \times T &\longrightarrow S \\ (a, b) &\longmapsto a. \end{aligned}$$

and

$$\begin{aligned} \text{proj}_2 : S \times T &\longrightarrow T \\ (a, b) &\longmapsto b. \end{aligned}$$

These definitions extend to sets S_1, \dots, S_n :

$$S_1 \cup \dots \cup S_n = \bigcup_{i \in I} S_i = \{a \mid a \in S_1 \text{ and } \dots \text{ and } a \in S_n\} \quad (1.1)$$

§1.1.1 The Well-Ordering Principle

Theorem 1.1.1: Well-Ordering Principle

Let $S \subset \mathbb{N}$ be a non-empty subset of \mathbb{N} . Then S has a *minimal element*. That is,

$\exists m \in S$ s.t. $n \geq m, \forall n \in S$. Informally, there exists a minimum element that is smaller than all other natural elements.

Proof. Since S is non-empty, we can pick $k \in S$. By definition of \mathbb{N} , we can start with 1 and add 1 repeatedly to get k . So, there are only k elements of \mathbb{N} less than or equal to k :

$$1 < 2 < \dots < k - 1 < k.$$

So, we can keep moving down from k , until we find an element $j \notin S$; since there are no smaller elements than $j + 1 \in S$, $j + 1$ is the minimal element. \square

§1.2 Functions

Definition 1.2.1: Functions

A **function** from S to T is a rule that assigns some element of T to each element of S :

$$f : S \rightarrow T, s \mapsto f(s).$$

S is the **domain**, and T the **codomain**.

Definition 1.2.2: Composition of Functions

If $f : S \rightarrow T$ and $g : T \rightarrow U$, then the **composition** of f and g is

$$g \circ f : S \rightarrow U, a \mapsto g(f(a)).$$

Definition 1.2.3: Bijectivity

Let $f : S \rightarrow T$ be a function.

1. f is **injective** or one-to-one if distinct elements of S go to distinct elements of T . In other words,

$$f(a) = f(b) \rightarrow a = b.$$

2. f is **surjective** or onto if every element of T comes from some element in S :

$$\forall t \in T, \exists s \in S \text{ s.t. } f(s) = t.$$

3. f is **bijective** if it is both injective and surjective.

Definition 1.2.4: Invertibility

Let $f : S \rightarrow T$ be a function. f is **invertible** if

$$\exists g : T \rightarrow S, (g \circ f)(s) = s, s \in S \text{ and } (f \circ g)(t) = t, t \in T.$$

Theorem 1.2.1: Bijective iff Invertible

Let $f : S \rightarrow T$ be a function. Then f is invertible $\iff f$ is bijective.

Proof. Suppose first that f is invertible. Let $g : T \rightarrow S$ denote the inverse. We need to prove that f is bijective.

To prove injectivity, suppose $f(a) = f(b)$ for some $a, b \in S$. Applying g to both sides and using the fact that g is the inverse of f , we have

$$g(f(a)) = g(f(b)) \Rightarrow a = b.$$

Thus f is injective.

To prove surjectivity, let $t \in T$; we need to find $s \in S$ such that $f(s) = t$. Using the inverse, let $s = g(t)$. Then

$$f(s) = f(g(t)) = t.$$

Thus f is surjective.

Since f is both injective and surjective, f is bijective.

Now, suppose that f is bijective. Then $\forall t \in T, \exists s \in S$ s.t. $f(s) = t$. Define a new function $g : T \rightarrow S$

$$g(t) := \text{"the unique } s \in S \text{ s.t. } f(s) = t\text{"}.$$

We now show that $(g \circ f)(s) = s$ and $(f \circ g)(t) = t$ for $s \in S, t \in T$.

Given $t \in T$, $f(g(t)) = t$ by definition of t . Given $s \in S$, we know that s maps to $f(s)$; so, by definition of g , $g(f(s)) = s$.

Thus, g is the inverse of f . □

Chapter 2

Groups: Part I

Groups are a fundamental baseline for abstract algebra. We start with motivating examples, then move on to a concrete definition.

§2.1 Motivation

§2.1.1 Permutations

Definition 2.1.1: Permutations

Let X be a set. A **permutation** of X is a bijective function

$$\pi : X \rightarrow X$$

with the property: $\forall x \in X, \exists x' \in X$ such that $\pi(x') = x$. This allows us to define an inverse π^{-1} to be the permutation

$$\pi^{-1} : X \rightarrow X$$

with the rule that $\pi^{-1}(x) = x'$, where $x' \in X$ is the unique element such that $\pi(x') = x$.

The **identity permutation** of X is the identity map

$$e : X \rightarrow X, e(x) = x, \forall x \in X.$$

In general, a *permutation* of a set X is a rule that “mixes up” the elements of X .

Example 2. Let $X = \{1, 2, 3, 4\}$. Then a permutation $\sigma : X \rightarrow X$ can be thought of as a *shuffling* of X and visualized as follows:

$$\begin{aligned} 1 &\Rightarrow 2 \\ 2 &\Rightarrow 3 \\ 3 &\Rightarrow 1 \\ 4 &\Rightarrow 4 \end{aligned}$$

σ^{-1} would be defined as

$$\begin{aligned} 1 &\Rightarrow 3 \\ 2 &\Rightarrow 1 \\ 3 &\Rightarrow 2 \\ 4 &\Rightarrow 4 \end{aligned}$$

Now, suppose τ is defined as $1 \Rightarrow 1, 2 \Rightarrow 3, 3 \Rightarrow 2, 4 \Rightarrow 4$. Then $\sigma \circ \tau$ is

$$\begin{aligned} 1 &\Rightarrow 2 \\ 2 &\Rightarrow 1 \\ 3 &\Rightarrow 3 \\ 4 &\Rightarrow 4 \end{aligned}$$

and $\tau \circ \sigma$ is

$$\begin{aligned} 1 &\Rightarrow 3 \\ 2 &\Rightarrow 2 \\ 3 &\Rightarrow 1 \\ 4 &\Rightarrow 4 \end{aligned}$$

From this, we gather some observations.

- Given any 2 permutations, we can compose to get a new one.
- There was a permutation that didn't do anything ($\sigma \circ \sigma^{-1}$).
- We can invert any permutation.
- If σ, τ are two permutations, then we don't necessarily have $\tau \circ \sigma = \sigma \circ \tau$ (in other words, the group of permutations with composition is not commutative).

Definition 2.1.2: Transformations

Let X be a figure in \mathbb{R}^2 . Then $Trafo(X)$ is the set of transformations on X .

Consider the symmetries of a square (involving reflections/rotations on a square) as a motivating example of transformations; are they invertible? commutative?

Remark 2. Each transformation gives a permutation of the vertices $\{A, B, C, D\}$.

§2.2 (Abstract) Groups

We now formally define the notion of a **group**.

Definition 2.2.1: Groups

A **group** $\{X, \cdot\}$ consists of a set X , together with a group rule/law

$$\begin{aligned} \cdot : G \times G &\rightarrow G \\ (g_1, g_2) &\mapsto g_1 \cdot g_2 \end{aligned}$$

satisfying the following axioms:

1. (identity) there is an element $e \in G$ such that

$$e \cdot g = g \cdot e = g.$$

for all $g \in G$.

2. (inverse) For all $g \in G$, there is an $h \in G$ such that

$$g \cdot h = h \cdot g = e.$$

The element h is called g^{-1} , the inverse of g .

3. (associativity) Given g_1, g_2, g_3 , we have

$$g_1(g_2 \cdot g_3) = (g_1 \cdot g_2)g_3.$$

If, in addition, the group satisfies

4. (commutative) Given $g_1, g_2 \in G$, we have

$$g_1 \cdot g_2 = g_2 \cdot g_1.$$

then G is an **Abelian** group.

Now, we observe some interesting properties that follow from the group axioms.

Proposition 2.2.1: Group Properties

Let G be a group.

1. The identity element is unique.
2. Each element of G has only one inverse.
3. If $g, h \in G$, then $(gh)^{-1} = h^{-1}g^{-1}$.
4. Given $g \in G$, $(g^{-1})^{-1} = g$.

Proof of (b). Suppose $g \in G$ and that both h_1, h_2 satisfy the inverse axiom. Then

$$g \cdot h_1 = e = g \cdot h_2.$$

By the inverse axiom, we multiply on the left by an inverse of g :

$$\begin{aligned} e \cdot h_1 &= e \cdot h_2 \\ h_1 &= h_2. \end{aligned}$$

Thus the inverse is unique. □

Definition 2.2.2: Order

- The **order** of a group G is denoted $\#G$ or $|G|$ is the number of elements in G if finite, and ∞ if infinite.
- If G is a group and $g \in G$, the smallest n in which $g^n = e$ is called **the order**

of g . If no n exists, we say g has infinite order.

Proposition 2.2.2: Individual Order and Group Order

Suppose G is a finite group and suppose $g^n = e$. Then the order of g divides n .

Proof. Let m be the order of $g \in G$; then m is the smallest positive integer such that $g^m = e$. Dividing n by m yields

$$n = mq + r, \quad q, r \in \mathbb{Z}, 0 \leq r < m.$$

In other words, dividing n by m leaves a quotient q and a remainder r . Using this equality together with $g^n = g^m = e$, we have

$$e = g^n = g^{mq+r} = (g^m)^q \cdot g^r = e^q \cdot g^r = g^r.$$

Hence $g^r = e$, and $r \in [0, m)$. But by definition, m is the smallest integer such that $g^m = e$. Therefore $r = 0$, and $n = mq$, and so m , the order of g , divides n . \square

Proposition 2.2.3: Order of Inverse

Let G be a finite group, and $g \in G$. Then $|g| = |g^{-1}|$.

Proof. Let $|g| = n$; then $g^n = e$. From this, we get

$$e = (g \cdot g^{-1})^n = g^n \cdot (g^{-1})^n = e \cdot (g^{-1})^n,$$

and so $(g^{-1})^n = e$.

Now we show that $|g^{-1}| = n$. Suppose $|g^{-1}| = m$, and $m < n$. Then

$$e = g^n \cdot (g^{-1})^m = g^{n-m}.$$

But we know that $|g| = n$, or equivalently, n is the smallest positive integer such that $g^n = e$; hence $g^{n-m} = e$ is a contradiction. Thus $m = n$, and so $|g^{-1}| = n$. \square

§2.2.1 Examples of Groups

Example 3. $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ and \mathbb{C} are all Abelian groups with respect to addition. However, \mathbb{Z} is not a group with respect to multiplication, as the multiplicative inverse does not exist. Additionally, \mathbb{Q}, \mathbb{R} , and \mathbb{C} are not groups with respect to multiplication, due to zero; but $\mathbb{Q} \setminus \{0\}, \mathbb{R} \setminus \{0\}$, and $\mathbb{C} \setminus \{0\}$ are all groups under multiplication.

Example 4. Let $\mathbb{Z}/m\mathbb{Z}$ be the set of integers modulo m . Then $\mathbb{Z}/m\mathbb{Z}$ is a group under addition modulo m , $+_m$; $\mathbb{Z}/m\mathbb{Z}$ is finite with order m . We also observe that $\mathbb{Z}/m\mathbb{Z}$ is a cyclic group.

Example 5. Let the set of $n \times n$ matrices be M_n . Then M_n is an Abelian group under addition, but not multiplication (since not all matrices have inverses).

Let

$$GL_n(\mathbb{R}) = \{M \in M_n \mid \det(M) \neq 0\}$$

denote the **general linear group**. Then $GL_n(\mathbb{R})$ is a non-Abelian group under matrix multiplication.

§2.2.2 Cyclic Groups

Definition 2.2.3: Cyclic Groups

A group G is **cyclic** if there is a $g \in G$ such that

$$G = \{\dots, g^{-2}, g^{-1}, e \text{ (or } g^0), g, g^2, g^3, \dots\}.$$

We call g a **generator**.

In general, for $n \geq 1$, the **abstract cyclic group order n** is the set

$$C_n = \{g_0, g_1, \dots, g_{n-1}\}$$

together with the composition rule

$$g_i \cdot g_j = \begin{cases} g_{i+j}, & i+j < n \\ g_{i+j-n}, & i+j \geq n \end{cases}$$

The identity element of C_n is g_0 , and the inverse of g_i is g_{n-i} (except g_0 , whose inverse is g_0). Further, C_n is an Abelian group, as $g_{i+j} = g_{j+i}$.

Some examples of cyclic groups are \mathbb{Z} and $\mathbb{Z}/m\mathbb{Z}$; both have generators 1. Another one is the permutation group.

Definition 2.2.4: Permutation Groups

Given X a set, let S_X denote the **symmetric group of X** , or the group of permutations of X . If

$$X = \{1, \dots, n\},$$

we use the notation S_n .

Let P_n be a regular n -gon with vertices $1, \dots, n$. The group of transformations of D_n (e.g. rotations, reflections, and compositions of such) is called the **dihedral group D_n** . We will later prove that D_n has order $2n$.

§2.3 Group Homomorphisms

Suppose that G, G' are groups, and suppose that ϕ is a function

$$\phi : G \longrightarrow G'$$

from elements of G to elements of G' . Many functions exist, but we're interested in the ones that preserve the "structure", or *group-iness*, of G and G' . But what makes a group a group? Specifically, groups are **associative**, and have **identity and inverse elements**. Thus, a function ϕ must preserve these qualities. We call such structure-preserving functions **homomorphisms**.

Definition 2.3.1: Homomorphisms

Let G_1, G_2 be groups. A **homomorphism** from G_1 to G_2 is a function

$$\phi : G_1 \rightarrow G_2$$

satisfying:

$$\phi(g_1 \cdot g_2) = \phi(g_1) \cdot \phi(g_2).$$

In other words, the map ϕ preserves the group operations. (Note that the composition law is different on the left and right sides! The left is the composition law of G_1 , while the right is the composition law of G_2 .)

It turns out this property is enough to force the identities and inverses to exist under a function.

Proposition 2.3.1

Let $\phi : G \rightarrow G'$ be a homomorphism of groups.

1. Let $e \in G$ be the identity element of G . Then $\phi(e) \in G'$ is the identity element of G' .
2. Let $g \in G$, and let $g^{-1} \in G$ be its inverse. Then $\phi(g^{-1}) \in G$ is the inverse of $\phi(g)$.

Proof. 1. Observe that $e = e \cdot e$, and that ϕ is a homomorphism (and so $\phi(e) = \phi(e \cdot e) = \phi(e) \cdot \phi(e)$). Let $e' \in G'$ be the identity element of G' . Then

$$\begin{aligned} e' &= \phi(e) \cdot \phi(e)^{-1} \\ &= (\phi(e) \cdot \phi(e)) \cdot \phi(e)^{-1} \\ &= \phi(e) \cdot (\phi(e) \cdot \phi(e)^{-1}) \\ &= \phi(e) \cdot e' \\ &= \phi(e). \end{aligned}$$

Hence $e' = \phi(e)$.

2. We have

$$\begin{aligned} \phi(g^{-1}) \cdot \phi(g) &= \phi(g^{-1} \cdot \phi(g)) \\ &= \phi(e) \\ &= e'. \end{aligned}$$

The proof that $\phi(g) \cdot \phi(g^{-1}) = e'$ is similar. Hence $\phi(g^{-1})$ is the inverse of $\phi(g)$. \square

Example 6. *Examples of homomorphisms:*

- There exists a homomorphism from the dihedral group to the group ± 1 :

$$\phi : D_n \rightarrow \{\pm 1\}$$

, where $\phi(\sigma) = 1$ if rotation, $\phi(\sigma) = -1$ if flip.

- For $n \geq m \geq 1$, there is an injective homomorphism

$$f : S_m \rightarrow S_n.$$

Note that this homomorphism is not surjective. More generally, if $X_1 \subseteq X_2$, then there is an injective homomorphism $f : S_{X_1} \rightarrow S_{X_2}$.

- There is a homomorphism

$$\log : (\mathbb{R}, \times) \rightarrow (\mathbb{R}, +).$$

- There is a homomorphism between the general linear group to the real numbers

$$\begin{aligned} \det : GL_n(\mathbb{R}) &\longrightarrow \mathbb{R} \\ AB &\longmapsto \det(AB) = \det(A) \cdot \det(B). \end{aligned}$$

Definition 2.3.2: Isomorphisms

Groups G_1, G_2 are **isomorphic** if there exists a **bijective homomorphism** $f : G_1 \rightarrow G_2$. In this case, f is called an **isomorphism**.

Interestingly, isomorphic groups are really the same group, but their elements are given different names.

We've now seen two examples of cyclic groups of order n : $\mathbb{Z}/n\mathbb{Z}$ and C_n . Naturally, we wonder if these groups are actually different (from the perspective of group theory). Equivalently, **are these two groups isomorphic?**

Example 7. $\mathbb{Z}/n\mathbb{Z}$ is isomorphic to C_n ($\mathbb{Z}/n\mathbb{Z} \cong C_n$). Consider the map

$$\begin{aligned} \phi : \mathbb{Z}/n\mathbb{Z} &\longrightarrow C_n \\ a &\longmapsto \phi(a) = g_a. \end{aligned}$$

Then $\phi(a+b) = \phi(a) \cdot \phi(b)$ by definition of group operations. So ϕ is a homomorphism. ϕ is surjective since $i \in \{0, \dots, n-1\}$ maps to $g_i \in \{g_0, \dots, g_{n-1}\}$. Since $\mathbb{Z}/n\mathbb{Z}$ and C_n both have n elements, ϕ is injective as well. So, ϕ is an isomorphism and $C_n \simeq \mathbb{Z}/n\mathbb{Z}$.

Note that if a group is isomorphic, there isn't necessarily a unique isomorphism. Consider the same isomorphism as above, except map $a \mapsto g_{a+1}$. This is also an isomorphism.

Example 8. Given any group G , and an element $g \in G$, then multiplication by g permutes the elements of G . This gives rise to an injective homomorphism $\phi : G \rightarrow S_G$.

This implies that by knowing every symmetric group, one knows much about every other group.

§2.4 Subgroups, Cosets, and Lagrange's Theorem

In all mathematics, a three-step process exists for studying complicated objects.

1. Deconstruction: Break your object into smaller and simpler pieces.
2. Analysis: Analyze the smaller, simpler pieces.
3. Fit the pieces back together.

For a group G , a natural way to form a smaller and simpler piece is by taking subsets $H \subseteq G$ that are themselves groups.

Definition 2.4.1: Subgroups

Let G be a group. A **subgroup of G** is a subset $H \subset G$ that is itself a group under G 's group law. Explicitly, H needs to satisfy

1. (Closure Under Composition) For every $h_1, h_2 \in H$, $h_1 \cdot h_2 \in H$
2. The identity element e is in H .
3. For every $h \in H$, its inverse h^{-1} is in H .

This is sometimes denoted $H < G$.

Note that since H uses G 's composition law, associativity is automatically satisfied. If H is finite, the **order** of H is the number of elements in H .

Proposition 2.4.1: Easier Subgroup Checking

Let G be a group, and $H \subseteq G$ a subset. If

- $H \neq \emptyset$
- For every $h_1, h_2 \in H$, the element $h_1 h_2^{-1}$ is in H

then H is a subgroup of G .

Proof. Clearly, $H \neq \emptyset$ (otherwise the identity would not be in H). To show that $e \in H$, let $h_2 = h_1$. Then

$$h_1 \cdot h_2^{-1} = h_1 \cdot h_1^{-1} = e \in H.$$

Thus the identity is in H .

To show that $\forall h \in H, h^{-1} \in H$, let $h_1 = e$. Then

$$h_1 \cdot h_2^{-1} = e \cdot h_2^{-1} = h_2^{-1} \in H.$$

Thus for any $h \in H$, its inverse h^{-1} is in H .

To show closure, observe that for any $h \in H, h^{-1} \in H$ (from above), and that $(h^{-1})^{-1} = h$. Let $h_2 = h^{-1}$. Then

$$h_1 \cdot h_2^{-1} = h_1 \cdot (h^{-1})^{-1} = h_1 \cdot h \in H.$$

Thus for any $h_1, h \in H$, we have $h_1 \cdot h \in H$. Thus H is closed.

Hence H is a subgroup of G . □

Example 9. Every group G has at least two subgroups, the **trivial subgroup** $\{e\}$ consisting of only the identity element, and the entire group G .

Example 10. Let G be a group, and let $g \in G$ be an element of order n . The **cyclic subgroup of G generated by g** , denoted $\langle g \rangle$, is the set

$$\langle g \rangle = \{\dots, g^{-2}, g^{-1}, e, g^1, g^2, g^3, \dots\}.$$

It is isomorphic to the cyclic group C_n .

If g has infinite order, then $\langle g \rangle \cong \mathbb{Z}$ ($\langle g \rangle$ is isomorphic to \mathbb{Z}).

Example 11. More examples of subgroups:

- Let $d \in \mathbb{Z}$; then we can form a subgroup of \mathbb{Z} using multiples of d , or $d\mathbb{Z}$.
- The set of rotations in the dihedral group \mathcal{D}_n is a subgroup of \mathcal{D}_n .

Every group homomorphism has an associated subgroup, the **kernel**, which can be a convenient check to see if the homomorphism is injective.

Definition 2.4.2: Kernel

Let $\phi : G \rightarrow G'$ be a group homomorphism. The **kernel of ϕ** , denoted $\ker(\phi)$, is the set of elements of G that are sent to the identity element of G' ,

$$\ker(\phi) = \{g \in G \mid \phi(g) = e'\}.$$

Example 12. *The kernel of the determinant homomorphism*

$$\det : \mathrm{GL}_n(\mathbb{R}) \longrightarrow \mathbb{R} \setminus \{0\}.$$

is

$$\ker(\det) = \{A \in \mathrm{GL}_n(\mathbb{R}) \mid \det(A) = 1\}.$$

We now observe two important properties of the kernel.

Proposition 2.4.2: Kernel Properties

Let $\phi : G \rightarrow G'$ be a group homomorphism.

1. $\ker(\phi)$ is a subgroup of G .
2. ϕ is injective if and only if $\ker(\phi) = \{e\}$.

Proof. We know that $\phi(e) = e'$, so $e \in \ker(\phi)$. Next, let $g_1, g_2 \in \ker(\phi)$. By the homomorphism property,

$$\phi(g_1 \cdot g_2) = \phi(g_1) \cdot \phi(g_2) = e' \cdot e' \in G',$$

so $g_1, g_2 \in \ker(\phi)$. Finally, for $g \in \ker(\phi)$, we know $\phi(g^{-1}) = \phi(g)^{-1} = e'^{-1} = e$, so $g^{-1} \in \ker(\phi)$. Thus, $\ker(\phi)$ is a subgroup of G .

Now, we know again that $e \in \ker(\phi)$ (since $\phi(e) = e'$). If ϕ is injective, by definition $\ker(\phi) = \{e\}$ (at most one element $g \in G$ satisfies $\phi(g) = e'$).

Now, suppose $\ker(\phi) = \{e\}$. Let $\phi(g_1) = \phi(g_2)$ for some $g_1, g_2 \in G$. Observe that $g_2^{-1} \in G$, and $\phi(g_2^{-1}) = \phi(g_2)^{-1}$. Then

$$\phi(g_1) = \phi(g_2) \implies \phi(g_1) \cdot \phi(g_2)^{-1} = \phi(g_2) \cdot \phi(g_2)^{-1} = e',$$

and so $\phi(g_1) \cdot \phi(g_2)^{-1} = \phi(g_1 \cdot g_2^{-1}) = e'$, which means $g_1 \cdot g_2^{-1} \in \ker(\phi) = \{e\}$. Hence $g_1 \cdot g_2^{-1} = e \implies g_1 = g_2$, and so ϕ is injective. \square

§2.4.1 Cosets

We can use a subgroup H of a group G to break G into pieces, called **cosets of H** .

Definition 2.4.3: Cosets

Let G be a group, and let $H < G$ be a subgroup. For each $g \in G$, the (left) **coset of H attached to g** is the set

$$gH = \{gh \mid h \in H\}.$$

In other words, gH is the resulting set we multiply g by every element $h \in H$.

Note that gH is **not** necessarily a subgroup of H ; sometimes $e \notin gH$.

We now prove several properties of cosets that help explain their importance.

Proposition 2.4.3: Properties of Cosets

Let G be a finite group, and let $H < G$.

1. Every element in G is in some coset of H .
2. Every coset of H has the same number of elements (namely, $|H|$).
3. Let $g_1, g_2 \in G$. Then the cosets g_1H and g_2H satisfy either

$$g_1H = g_2H \text{ or } g_1H \cap g_2H = \emptyset.$$

In other words, g_1H and g_2H are either equal or disjoint.

Proof. 1. Let $g \in G$. Since $e \in H$ for any subgroup $H < G$, the coset gH contains $g \cdot e = g$.

2. Let $g \in G$. To prove that the cosets gH and H have the same number of elements, we show the map

$$\begin{aligned} F : H &\longrightarrow gH \\ h &\longmapsto F(h) = gh \end{aligned}$$

is a bijective map from H to gH .

We first check that F is injective. Suppose $h_1, h_2 \in H$ satisfy $F(h_1) = F(h_2)$. Then $gh_1 = gh_2$, and multiplying by g^{-1} , we get $h_1 = h_2$. Hence F is injective. For surjectivity, observe that every element of gH looks like gh for some $h \in H$, and $F(h) = gh$, so every element of gH is the image of some element of H . Hence F is surjective.

Thus F is bijective, so H and gH have the same number of elements. Since this is true for any $g \in G$, every coset of H has the same number of elements.

3. If $g_1H \cap g_2H = \emptyset$, we are done, so assume the two cosets are not disjoint. Then there are some elements $h_1, h_2 \in H$ such that $g_1h_1 = g_2h_2$. Since $h_1^{-1} \in H$, we rewrite this as $g_1 = g_2h_2h_1^{-1}$. Now, take any element $a \in g_1H$. a is of the form g_1h for some $h \in H$. Then

$$a = g_1h = g_2h_2h_1^{-1}h \in g_2H,$$

as H is a subgroup, so $h_2h_1^{-1}h \in H$. Hence $g_1H \subseteq g_2H$; and from above, every coset has the same number of elements, so $g_1H \subseteq g_2H \implies g_1H = g_2H$. □

These properties lead to a fundamental divisibility property for the orders of subgroups.

Theorem 2.4.1: Lagranges Theorem

Let G be a finite group, and let $H < G$. Then the order of H divides the order of G ; or, $|G| = k|H|$, $k \in \mathbb{Z}$.

Proof. We start by choosing $g_1, \dots, g_k \in G$ so that g_1H, \dots, g_kH is a list of every different coset of H . Since every element of G is in some coset of H , we have that G is equal to the union of the cosets of H , namely

$$G = g_1H \cup \dots \cup g_kH.$$

Additionally, we know that distinct cosets share no elements, so if $i \neq j$, then $g_i H \cap g_j H = \emptyset$. Thus the union of cosets is a disjoint union, so the number of elements in G is the sum of the number of elements in each coset:

$$|G| = |g_1 H| + \dots + |g_k H|.$$

But we know that every coset of H has the same number of elements, so $|g_i H| = |H|$. Thus, we get

$$|G| = k |H|.$$

Thus the order of G is a multiple of the order of H . \square

Definition 2.4.4: Index

Let G be a group, and $H < G$. The **index of H in G** , denoted $(G : H)$, is the number of distinct cosets of H . In Lagrange's Theorem, the index $(G : H) = k$; so

$$|G| = (G : H) |H|.$$

Corollary 2.4.1: Extension of Lagrange's Theorem to Finite Groups

Let G be a finite group, and let $g \in G$. Then the order of g divides the order of G .

Proof. The order of the subgroup $\langle g \rangle$ generated by G is equal to the order of the element g , and Lagrange's Theorem tells us that the order of $\langle g \rangle$ divides the order of G . \square

We now give one application of Lagrange's Theorem, which marks the beginning of a long and ongoing mathematical journey that strives to classify finite groups according to their orders.

Proposition 2.4.4: Prime-Ordered Groups

Let p be a prime, and let G be a finite group of order p . Then G is isomorphic to the cyclic group \mathcal{C}_p .

Proof. Since $p \geq 2$, we know that G contains more than just the identity element, so we choose some non-identity element $g \in G$.

By Lagrange's Theorem, we know that the order of the subgroup $\langle g \rangle$ divides the order of G . But since $|G| = p$ is prime, the order of $\langle g \rangle$ is either 1 or p ; and since $\langle g \rangle$ contains both e and g (and so $\langle g \rangle > 1$), we know $|\langle g \rangle| = p = |G|$. Thus the subgroup $\langle g \rangle$ has the same number of elements as the full group, so they are equal: $\langle g \rangle = G$.

Now, we denote the cyclic group $\mathcal{C}_p = \{g_0, g_1, \dots, g_{p-1}\}$. We obtain an isomorphism

$$\begin{aligned} \mathcal{C}_p &\longrightarrow G \\ g_i &\longmapsto g^i. \end{aligned}$$

Thus G is isomorphic to \mathcal{C}_p . \square

Remark 3. *The vast theory of finite groups has many fascinating (and frequently unexpected) results, with easy to understand statements, yet surprisingly intricate proofs. Two such theorems are stated.*

Theorem 2.4.2

Let p be a prime number, and let G be a group of order p^2 . Then G is an Abelian group.

On the other hand, we know that there exist non-Abelian groups of order p^3 . For instance, \mathcal{D}_4 and the quaternion group \mathcal{Q} are non-Abelian groups of order $8 = 2^3$.

The next result is a partial converse of Lagrange's Theorem.

Theorem 2.4.3: Sylow's Theorem

Let G be a finite group, let p be a prime, and suppose p^n divides $|G|$ for some power $n \geq 1$. Then G has a subgroup of order p^n .

One might hope, more generally, that if d is any number that divides the order of G , then G has a subgroup of order d . Unfortunately, this is not true; however, we have not yet seen a counterexample.

Both theorems will be proved later.

§2.5 Products of Groups

Subgroups provide a way to break complicated objects (groups) down into smaller, simpler pieces. We now look at a way in which two smaller groups can be used to build a larger group.

Definition 2.5.1: Products of Groups

Let G_1, G_2 be groups. The **product** of G_1 and G_2 is the group whose elements consist of ordered pairs

$$G_1 \times G_2 = \{(a, b) \mid a \in G_1 \text{ and } b \in G_2\},$$

and whose group operation is performed separately on each component. In other words, if the group operation of $G_1 \times G_2$ is $*$, the group operation of G_1 is \cdot , and the group operation of G_2 is \circ , we have

$$(a_1, b_1) * (a_2, b_2) = (a_1 \cdot a_2, b_1 \circ b_2).$$

It is clear that the identity element of $G_1 \times G_2$ is (e_1, e_2) , and the inverse of an element $(a, b) \in G$ is given by

$$(a^{-1}, b^{-1}).$$

More generally, we can take any list of groups G_1, \dots, G_n and form the product

group

$$G_1 \times \dots \times G_n.$$

Remark 4. We observe that $G = G_1 \times G_2$ has order $|G_1| \cdot |G_2|$.

Example 13. For any non-zero numbers m, n , there is a homomorphism

$$\begin{aligned} \mathbb{Z}/mn\mathbb{Z} &\longrightarrow (\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z}) \\ a \bmod (mn) &\longmapsto (a \bmod m, a \bmod n). \end{aligned}$$

We claim that if $\gcd(m, n) = 1$, then it is an isomorphism. To see this, suppose that $a \bmod mn$ is in the kernel. Then

$$a \equiv 0 \bmod m \text{ and } a \equiv 0 \bmod n.$$

In other words, a is divisible by both m and n , and then the assumption that $\gcd(m, n) = 1$ implies that a is divisible by mn . Thus $a \equiv 0 \bmod mn$, which proves that the kernel of the homomorphism is $\{0\}$ (and thus the homomorphism is injective). Further, since the finite set $\mathbb{Z}/mn\mathbb{Z}$ has the same number of elements as $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ ($\mathbb{Z}/mn\mathbb{Z}$ has mn elements, while $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ has $m \cdot n = mn$ elements), so it is surjective, and thus the homomorphism is an isomorphism.

One interpretation of this example is that it tells us that if $\gcd(m, n) = 1$, then the large group $\mathbb{Z}/mn\mathbb{Z}$ may be broken down into the product of two smaller groups $\mathbb{Z}/m\mathbb{Z}$ and $\mathbb{Z}/n\mathbb{Z}$. Repeated applications demonstrate that any finite cyclic group is isomorphic to the product of cyclic groups of prime power order. The following theorem extends this to all finite Abelian groups.

Theorem 2.5.1: Structure Theorem for Finite Abelian Groups

Let G be a finite Abelian group. Then there are integers m_1, \dots, m_r so that

$$G \cong (\mathbb{Z}/m_1\mathbb{Z}) \times (\mathbb{Z}/m_2\mathbb{Z}) \times \dots \times (\mathbb{Z}/m_r\mathbb{Z}).$$

Example 14. (Projects and Inclusions) Products of groups come with two natural *projection homomorphisms*

$$\begin{aligned} p_1 : G_1 \times G_2 &\longrightarrow G_1, & p_2 : G_1 \times G_2 &\longrightarrow G_2, \\ (a, b) &\longmapsto a, & (a, b) &\longmapsto b, \end{aligned}$$

and two natural *inclusion homomorphisms*

$$\begin{aligned} \iota_1 : G_1 &\longrightarrow G_1 \times G_2 & \iota_2 : G_2 &\longrightarrow G_1 \times G_2 \\ a &\longmapsto (a, e_2) & b &\longmapsto (e_1, b). \end{aligned}$$

The inclusion maps are clearly injective, but the projections have kernels

$$\ker(p_1) = \{e_1\} \times G_2 \text{ and } \ker(p_2) = G_1 \times \{e_2\}.$$

Chapter 3

Rings: Part I

Unlike groups, which were completely new, the concept of a **ring** is mildly familiar! Some examples of rings include:

- \mathbb{Z} , \mathbb{Q} , \mathbb{R} , and \mathbb{C} are rings (\mathbb{Q} , \mathbb{R} , and \mathbb{C} are actually **fields**, a special type of ring; such is a discussion for later)
- The set of integers modulo m is a ring

These examples all share something in common: they each have two operations, "addition" and "multiplication", and each operation individually satisfies some axioms, along with the great and powerful distributive law.

In general, a ring is a set with two operations satisfying a bunch of axioms that are modeled after the properties of addition and multiplication of integers. We will later formalize this; but first, a little number theory.

§3.1 Review of Number Theory

§3.1.1 Equivalence Relations

We first introduce the notion of **equivalence relations**; while not strictly related to number theory, equivalence relations will be significant for modular arithmetic.

Definition 3.1.1: Equivalence Relations

An **equivalence relation** on a set S is a relation " \sim " satisfying

1. **Reflexivity**: For $a \in S$, $a \sim a$
2. **Symmetry**: For $a, b \in S$, $a \sim b$ implies $b \sim a$
3. **Transitivity**: For $a, b, c \in S$, if $a \sim b$ and $b \sim c$, then $a \sim c$.

$a \sim b$ means a is "related" to b ; $a \not\sim b$ means a is "not related" to b .

Given an $a \in S$, the **equivalence class** of a is

$$S_a = \{b \in S \mid b \sim a\}.$$

Note that S_a is never empty; it always contains a .

Some examples of equivalence relations are equality ($=$) and congruence \pmod{m} ; on the other hand, order (e.g. \leq) is **not** an equivalence relation (symmetry does not hold). We now look further into the congruence \pmod{m} equivalence relation.

Example 15. Given $a \in \mathbb{Z}$, $b \equiv a \pmod{m}$ iff

$$n|b-a \iff b-a=kn, \ k \in \mathbb{Z} \iff b=a+kn.$$

So \mathbb{Z}_a actually forms a coset of $n\mathbb{Z}$:

$$\begin{aligned} \mathbb{Z}_a &= \{b \in \mathbb{Z} \mid b \equiv a \pmod{m}\} \\ &= \{a + kn \mid k \in \mathbb{Z}\} \\ &= a + n\mathbb{Z}. \end{aligned}$$

That is, each equivalence class for congruence \pmod{m} is actually a coset of $m\mathbb{Z}$ in \mathbb{Z} :

$$\mathbb{Z}/m\mathbb{Z} = \text{set of cosets of } m\mathbb{Z} \text{ in } \mathbb{Z}.$$

Theorem 3.1.1

Let S be a set with an equivalence relation \sim . Then

1. If $a, b \in S$, then either

$$S_a \cap S_b = \emptyset \text{ or } S_a = S_b.$$

2. Let $\{C_i\}_{i \in I}$ be the disjoint equivalence classes of S . Then

$$S = \coprod_{i \in I} C_i.$$

In particular, if S is finite, then

$$|S| = |C_1| + \dots + |C_n|.$$

§3.1.2 Modular Arithmetic

We first observe an important characteristic of the gcd:

Proposition 3.1.1

Given integers u, v , there exists integers x, y such that

$$ux + vy = \gcd(u, v).$$

x, y can be found using the **Euclidean algorithm**.

This result leads us to an important proposition that facilitates equivalence relations in $\mathbb{Z}/m\mathbb{Z}$:

Proposition 3.1.2

$ax \equiv b \pmod{m}$ is solvable if and only if

$$\gcd(a, m) \mid b.$$

Proof. First, suppose $ax \equiv b \pmod{m}$ is solvable. It follows that

$$m \mid ax - b.$$

So, we can find $k \in \mathbb{Z}$ such that $ax - b = km$ is $b = ax - km$. From this, we get that any integer that divides both a and m must divide b ; in particular, $\gcd(a, m) \mid b$.

Conversely, if $\gcd(a, m) \mid b$, then $b = c \gcd(a, m)$ for some $c \in \mathbb{Z}$. Using the Euclidean algorithm to find x, y with

$$ax + ym = \gcd(a, m),$$

and multiplying by c to get

$$a(cx) + y(mc) = c \gcd(a, m) = b,$$

we have $a(cx) \equiv b \pmod{m}$, so the congruence is solvable. \square

One natural followup is:

When does $ax \equiv 1 \pmod{m}$ have a solution?

From the proposition above, **only if** $\gcd(a, m) = 1$; that is, only if a and m are relatively prime. Rephrasing, if a and m are relatively prime, then a has a unique multiplicative inverse in $\mathbb{Z}/m\mathbb{Z}$.

§3.2 Abstract Rings and Ring Homomorphisms

Definition 3.2.1: Rings

A **ring** R is a set with two operations, generally called **addition** and **multiplication** and written

$$\underbrace{a + b}_{\text{addition}}$$

and

$$\underbrace{a \cdot b \text{ or } ab}_{\text{multiplication}}$$

satisfying the following axioms:

1. The set R with addition law $+$ is an Abelian group, with identity 0 (or 0_R).
2. The set R with multiplication law \cdot is a **monoid** (associative, identity, **but no inverse**), with identity 1 (or 1_R ^a).
3. **Distributive Law:** For all $a, b, c \in R$, we have

$$a \cdot (b + c) = a \cdot b + a \cdot c \quad \text{and} \quad (b + c) \cdot a = b \cdot a + c \cdot a.$$

4. If, in addition to these three properties, $a \cdot b = b \cdot a$ for all $a, b \in R$, then R is a **commutative ring**.

^ato avoid the trivial ring, we require $1_R \neq 0_R$; however, this is not strictly required

Experience with the integers seems to suggest that $0_R \cdot a = 0_R$, and $(-a) \cdot (-b) = a \cdot b$; yet why are these true? 0_R is the definition of the identity element for *addition*, so why should it say anything about *multiplication*? Similarly, $-a$ relates to the definition of *additive* inverse, but what does that tell us about its product with other elements in R ? To show these intuitively obvious claims, we need the distributive law.

Proposition 3.2.1

1. $0_R \cdot a = 0_R$ for all $a \in R$.
2. $(-a) \cdot (-b) = a \cdot b$ for all $a, b \in R$. In particular, we have $(-1_R) \cdot a = -a$.

Proof. 1. Note that $1_R = 0_R + 0_R$. Then

$$\begin{aligned}
 a &= 1_R \cdot a && [1_R \text{ is the multiplicative identity}] \\
 &= (1_R + 0_R) \cdot a && [\text{from above}] \\
 &= 1_R \cdot a + 0_R \cdot a && [\text{from distributivity}] \\
 &= a + 0_R \cdot a.
 \end{aligned}$$

Adding $-a$ to both sides, we get

$$0_R = 0_R + 0_R \cdot a,$$

and so $0_R \cdot a = 0_R$.

2. First, we show that $(-1_R) \cdot a = -a$:

$$\begin{aligned}
 a + (-1_R) \cdot a &= 1_R \cdot a + (-1_R) \cdot a \\
 &= (1_R + (-1_R)) \cdot a \\
 &= 0_R \cdot a \\
 &= 0_R.
 \end{aligned}$$

Hence $(-1_R) \cdot a$ is the inverse of a , and so $-a = (-1_R) \cdot a$.

Now, observe that $-ab = (-a)b$ (this proof is left as an exercise for the reader). Then

$$\begin{aligned}
 (-a) \cdot (-b) + -ab &= (-a) \cdot (-b) + (-a) \cdot b \\
 &= (-a) \cdot (-b + b) \\
 &= (-a) \cdot 0_R \\
 &= 0_R.
 \end{aligned}$$

Thus $(-a) \cdot (-b)$ is the inverse of $-ab$, and so $(-a) \cdot (-b) = ab$. □

Just like groups, we want to investigate maps

$$\phi : R \rightarrow R'$$

from one ring to another that respect the *ring-iness* of R and R' . Since rings are characterized by their addition and multiplication properties, we get the following definition.

Definition 3.2.2: Ring Homomorphisms

Let R, R' be rings. A **ring homomorphism** from R to R' is a function $\phi : R \rightarrow R'$ satisfying^a

1. $\phi(1_R) = 1_{R'}$
2. $\phi(a + b) = \phi(a) + \phi(b)$ for all $a, b \in R$
3. $\phi(a \cdot b) = \phi(a) \cdot \phi(b)$ for all $a, b \in R$

The **kernel** of ϕ is the set of elements that are sent to 0:

$$\ker(\phi) = \{a \in R \mid \phi(a) = 0_{R'}\}.$$

As with groups, R and R' are **isomorphic** if there is a bijective ring homomorphism $\phi : R \rightarrow R'$, and we call such a map an **isomorphism**.

^awe have the first axiom to disallow the boring and trivial zero map $\phi : R \rightarrow R', \phi(a) = 0_{R'}$.

§3.3 Interesting Examples of Rings

As described before, we have the four rings

$$\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C};$$

we say that \mathbb{Z} is a **subring** of \mathbb{Q} , and similarly for the others. Another example is the ring of integers modulo m , $\mathbb{Z}/m\mathbb{Z}$.

Example 16. (*Integers Modulo m $\mathbb{Z}/m\mathbb{Z}$*) We construct a ring $\mathbb{Z}/m\mathbb{Z}$ with integers, and determining equality if $a - b$ is a multiple of m . Formally, we define an equivalence relation on \mathbb{Z} by the rule

$$a \equiv b \text{ if } a - b = km \text{ for some } k \in \mathbb{Z}. \text{ We say } a \text{ is } \textbf{congruent} \text{ to } b \text{ modulo } m.$$

While $\mathbb{Z}/m\mathbb{Z}$ is not a subring of \mathbb{C} , there is a natural homomorphism

$$\phi : \mathbb{Z} \longrightarrow \mathbb{Z}/m\mathbb{Z}, \phi(a) = a \pmod{m}$$

that assigns the integer a to its equivalence class of all integers congruent to $a \pmod{m}$. This homomorphism ϕ is called the **reduction mod m homomorphism**. The kernel of ϕ is the set of all multiples of m .

Example 17. Another subring of \mathbb{C} is the **ring of Gaussian integers** $\mathbb{Z}[i]$, which consists of

$$\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\}.$$

The quantity i represents the imaginary number (e.g. $\sqrt{-1} = i$), and addition and multiplication are defined following the usual rules of adding and multiplying complex numbers.

In general, we can define a ring

$$\mathbb{F}[z] = \{a + bz \mid a, b \in \mathbb{F}\}$$

for an arbitrary field and any number z .

Example 18. Polynomials offer another method of creating bigger rings from already known rings. For any commutative ring R , we use R to build the **ring of polynomials over R** ,

$$R[x] = \{ \text{polynomials } a_0 + a_1x + \dots + a_dx^d \text{ of all possible degrees with coefficients } a_0, a_1, \dots, a_d \in R \}.$$

A common one we've seen before is $\mathbb{R}[x]$, but the rules of adding and multiplying polynomials hold in any commutative ring. Indeed, the rule for multiplying polynomials is a result of the distributive law!

Now, consider the polynomial

$$f(x) = a_0 + a_1x + \dots + a_dx^d \in R[x].$$

Then for any element $c \in R$, we can **evaluate f at c** simply by substituting x with c :

$$f(c) = a_0 + a_1c + \dots + a_dc^d \in R.$$

When we first studied polynomials, we viewed them as functions, i.e. $f(x)$ defined a function $f : R \rightarrow R$. While these polynomial functions are interesting, they are almost never ring homomorphisms!

Thus, we take another approach: using a particular $c \in R$, we define a function from the ring of polynomials $R[x]$ to the ring R :

$$E_c : R[x] \longrightarrow R, \quad E_c(f) = f(c).$$

We call E_c the **evaluation at c map**. If R is commutative, then E_c is a ring homomorphism, and its kernel is the set of polynomials that have a factor of $x - c$ (since $a_0 + a_1(c - c) + \dots + a_d(c - c)^d = 0_R$).

Example 19. One famous non-commutative ring is the **ring of quaternions**,

$$\mathbb{H} = \{a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} \mid a, b, c, d \in \mathbb{R}\}.$$

\mathbf{i} , \mathbf{j} , and \mathbf{k} are the different square roots of -1 , and although they commute with elements in \mathbb{R} , they don't commute with each other. To multiply two quaternions, we first use the distributive law, and then apply multiplication as specified in the quaternion group. Since \mathcal{Q} is a non-commutative group, the ring of quaternions \mathbb{H} is a non-commutative ring.

The ring of quaternions played an important role in the development of math and physics because of the **cancellation law**:

if you know that α and β are real (or complex) numbers satisfying $\alpha\beta = 0$, then either $\alpha = 0$ or $\beta = 0$. It turns out that the same is true with quaternions!

Example 20. (*Matrix Rings*) There are also rings with matrix elements. Let

$$M_2(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid a, b, c, d \in \mathbb{R} \right\}$$

denote the set of 2×2 matrices with real entries. Matrices are added by their corresponding elements, and multiplied using matrix multiplication. With these operations, $M_2(\mathbb{R})$ is a non-commutative ring. However, $M_2(\mathbb{R})$ does not satisfy the cancellation law:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

More generally, the set of $n \times n$ matrices over a ring R , $M_n(R)$, forms a non-commutative ring.

There are many interesting homomorphisms from rings to matrix rings. For example,

$$\mathbb{C} \hookrightarrow M_2(\mathbb{R}), \quad x + yi \mapsto \begin{pmatrix} x & y \\ -y & x \end{pmatrix}$$

is an injective ring homomorphism (proof left as an exercise).

Example 21. For every ring R , there is a unique homomorphism

$$\phi : \mathbb{Z} \longrightarrow R.$$

To understand why, note that by homomorphism requirements we must have $\phi(1) = 1_R$, and we must also have

$$\phi(n) = \phi(1 + \dots + 1) = \phi(1) + \dots + \phi(1).$$

But we also need $\phi(0) = 0_R$, and $\phi(-n) = -\phi(n)$, so there are really no other choices for ϕ . In other words, the requirements that

$$\phi(1) = 1_R \text{ and } \phi : \mathbb{Z} \longrightarrow R \text{ is a homomorphism}$$

means that there is only one possibility for ϕ . One still needs to check that ϕ is a homomorphism; such a proof is left as an exercise.

§3.4 Important Properties of Rings

Some rings, such as \mathbb{Q} , \mathbb{R} , \mathbb{C} have the special property that every non-zero element has a multiplicative inverse. These rings are special, and we call them **fields**.

Definition 3.4.1: Fields

A **field** is a commutative ring R with the property that every non-zero element of R has a multiplicative inverse. In other words, for every non-zero $a \in R$, there is a $b \in R$ satisfying $ab = 1$.

Example 22. In addition to \mathbb{Z} , \mathbb{R} , and \mathbb{C} , there are also **finite fields**. One important example of a finite field is the ring $\mathbb{Z}/p\mathbb{Z}$, where p is a prime. This follows from number theory: if $p \nmid a$, then $\gcd(p, a) = 1$, and so $ax \equiv 1 \pmod{p}$ is solvable, e.g. there is a $b \in \mathbb{Z}/p\mathbb{Z}$ with $ab \equiv 1 \pmod{p}$. We denote this field \mathbb{F}_p .

Many other rings are not fields, such as \mathbb{Z} , $\mathbb{Z}[i]$, and $\mathbb{R}[x]$; however, they do have the nice property of cancellation:

Definition 3.4.2: Cancellation Property

Let R be a commutative ring. R has the **cancellation property** if for every $a, b, c \in R$,

$$ab = ac \text{ with } a \neq 0 \iff b = c.$$

Rings that maintain the cancellation property are called **integral domains**:

Definition 3.4.3: Zero Divisors and Integral Domains

Let R be a ring. An element $a \in R$ is a **zero divisor** if $a \neq 0$ and there is some non-zero $b \in R$ such that $ab = 0$. The ring R is an **integral domain** if it has no zero divisors. Equivalently, the ring R is an integral domain if the only way to get $ab = 0$ is if either $a = 0$ or $b = 0$.

In fact, every field is an integral domain, and a ring R is an integral domain if and only if it has the cancellation property. Moreover, every integral domain is a subring of a field (the reader should verify all of these statements). We will see later that the smallest such field is called the **field of fractions over R** .

§3.5 Unit Groups and Product Rings

In groups, we saw that many interesting subgroups and larger groups could be formed from a group. Similarly, we get the notion in rings that every ring contains an interesting group, and smaller rings can form larger rings.

§3.5.1 Unit Groups

Definition 3.5.1: Unit Groups

Let R be a commutative ring^a. The **group of units of R** is the subset R^* of R defined by

$$R^* = \{a \in R \mid \text{there exists some } b \in R \text{ satisfying } ab = 1\},$$

where group law is ring multiplication. Elements of R^* are called **units**.

^aFor a non-commutative ring R , $a \in R$ is a **unit** if there are elements $b, c \in R$ such that $ab = ca = 1$, i.e. the element a needs both a left- and right-inverse.

Proposition 3.5.1

The set of units R^* of R is a group, with group law being ring multiplication.

Proof. For each of the group axioms:

- Let $a_1, a_2 \in R^*$, and let $b_1, b_2 \in R^*$ be values such that $a_1 b_1 = 1$, $a_2 b_2 = 1$ (we can say $b_1, b_2 \in R^*$, not just R , since commutativity ensures that if $a \in R^*$, $b \in R^*$ as well). Then

$$\begin{aligned} 1 &= a_1 b_1 a_2 b_2 \\ &= a_1 a_2 b_1 b_2, \end{aligned}$$

and since $b_1 b_2 \in R^*$ (multiplication is closed in monoids), we have $a_1 a_2 \in R^*$.

- $1 \in R$ is the identity element.
- By definition, units have inverses.
- Multiplicative associativity is guaranteed by the ring axioms.

Hence R^* is a group. □

Example 23. Some unit groups include

$$\mathbb{Z}^* = \{\pm 1\}, \quad \mathbb{Z}[i]^* = \{\pm 1, \pm i\}, \quad \mathbb{R}[x]^* = \mathbb{R}^*.$$

Another interesting example is the ring $\mathbb{Z}[\sqrt{2}]$, whose unit group has infinitely many elements. The proofs of these assertions is left as an exercise for the reader.

Example 24. A ring R is a field if and only if

$$R^* = \{a \in R \mid a \neq 0\} = R \setminus \{0\};$$

in other words, every non-zero element has a multiplicative inverse.

We now explore the unit group of the ring $\mathbb{Z}/m\mathbb{Z}$.

Proposition 3.5.2

Let $m \geq 1$ be an integer. Then

$$(\mathbb{Z}/m\mathbb{Z})^* = \{a \bmod m \mid \gcd(a, m) = 1\}.$$

In particular if m is a prime number, then $\mathbb{Z}/p\mathbb{Z}$ is a field (often denoted \mathbb{F}_p).

Proof. Suppose that $\gcd(a, m) = 1$. Then by the Euclidean algorithm, we can find $u, v \in \mathbb{Z}$ satisfying $au + mv = 1$. Hence

$$au = 1 - mv \equiv 1 \pmod{m}.$$

Thus u is a multiplicative inverse for a in the ring $\mathbb{Z}/m\mathbb{Z}$, so $a \pmod{m}$ is in $(\mathbb{Z}/m\mathbb{Z})^*$. In the other direction, suppose that $a \pmod{m} \in (\mathbb{Z}/m\mathbb{Z})^*$. Then for any $a \in \mathbb{Z}/m\mathbb{Z}$, we can find some $b \pmod{m} \in \mathbb{Z}/m\mathbb{Z}$ such that

$$(a \pmod{m})(b \pmod{m}) = 1 \pmod{m}.$$

In other words, $ab \equiv 1 \pmod{m}$, so $ab - 1 = cm$ for some c . But then $ab - cm = 1$, and so $\gcd(a, m) = 1$, since any number dividing both a, m divides 1, which is only true for 1. \square

§3.5.2 Product Rings

Now, we inspect building larger rings from smaller rings. Why make things more complicated? It turns out that reversing this process, breaking up complicated rings into smaller, easier rings, can be useful; one such example is the Chinese Remainder Theorem. The building procedure is analogous to building products of groups, as well as constructing vector spaces (e.g. \mathbb{R}^n , making n -tuples from \mathbb{R}).

Definition 3.5.2: Products of Rings

Let R_1, \dots, R_n be rings. The **product of** R_1, \dots, R_n is the ring

$$R_1 \times \dots \times R_n = \{(a_1, \dots, a_n) \mid a_1 \in R_1, \dots, a_n \in R_n\}.$$

In other words, the product $R_1 \times \dots \times R_n$ is the set of n -tuples, where the first entry is chosen from R_1 , second from R_2 , etc. $R_1 \times \dots \times R_n$ becomes a ring using coordinate-wise addition and multiplication:

$$\begin{aligned} (a_1, \dots, a_n) + (b_1, \dots, b_n) &= (a_1 + b_1, \dots, a_n + b_n) \\ (a_1, \dots, a_n) \cdot (b_1, \dots, b_n) &= (a_1 \cdot b_1, \dots, a_n \cdot b_n). \end{aligned}$$

Proving $R_1 \times \dots \times R_n$ is a group is left as an exercise for the reader.

Example 25. The product ring $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$ has 6 elements,

$$(0, 0), (1, 0), (0, 1), (0, 2), (1, 1), (1, 2).$$

For example, addition and multiplication look like

$$(1, 1) + (1, 2) = (0, 0), (0, 2) \cdot (1, 2) = (0, 1).$$

It turns out that $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z} \cong \mathbb{Z}/6\mathbb{Z}$.

However, $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$ is **not** isomorphic to $\mathbb{Z}/8\mathbb{Z}$. To see this, if

$$\phi : \mathbb{Z}/8\mathbb{Z} \longrightarrow \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$$

is a homomorphism, then by definition $\phi(1) = (1, 1)$, so

$$\phi(4) = \phi(1+1+1+1) = \phi(1) + \phi(1) + \phi(1) + \phi(1) = (1, 1) + (1, 1) + (1, 1) + (1, 1) = (0, 0).$$

Hence $\ker(\phi)$ is non-trivial, so ϕ cannot be injective.

Now, we combine product rings and unit groups.

Proposition 3.5.3

Let R_1, \dots, R_n be rings. Then the unit groups of the product is isomorphic to the product of the unit groups:

$$(R_1 \times \dots \times R_n)^* \cong R_1^* \times \dots \times R_n^*.$$

Proof. If $(a_1, \dots, a_n) \in (R_1 \times \dots \times R_n)^*$, then by definition there is a $(b_1, \dots, b_n) \in (R_1 \times \dots \times R_n)^*$ satisfying

$$(a_1, \dots, a_n) \cdot (b_1, \dots, b_n) = (1, \dots, 1).$$

However, this means that $a_i b_i = 1$, so $a_i \in R_i^*$. Hence $(a_1, \dots, a_n) \in R_1^* \times \dots \times R_n^*$.

Now, suppose $(a_1, \dots, a_n) \in R_1^* \times \dots \times R_n^*$. Then for $a_i \in R_i^*$, there exists some $b_i \in R_i^*$ such that $a_i b_i = 1$. Then

$$(a_1, \dots, a_n) \cdot (b_1, \dots, b_n) = (1, \dots, 1).$$

Hence $(a_1, \dots, a_n) \in (R_1 \times \dots \times R_n)^*$. □

§3.6 Ideals and Quotient Rings

Recall that in the ring $\mathbb{Z}/m\mathbb{Z}$, we pretend that $a, b \in \mathbb{Z}$ are "identical" if $a - b = km$ for some $k \in \mathbb{Z}$. In other words, we get an equivalence relation

$$a \equiv b \pmod{m} \text{ if } a - b = km \text{ (} a - b \text{ is a multiple of } m \text{)}.$$

We then defined $\mathbb{Z}/m\mathbb{Z}$ to be the set of equivalence classes.

Now, we attempt to generalize this construction to arbitrary (commutative) rings. We first start by generalizing the concept "being a multiple of m ". Why would we want this?

- First, it provides us with a new ring to explore; when is $\mathbb{Z}/m\mathbb{Z}$ a field? An integral domain? Moreover, when exploring conjectures, $\mathbb{Z}/m\mathbb{Z}$ provides a testing ground.
- Second, sometimes $\mathbb{Z}/m\mathbb{Z}$ is easier to work with than \mathbb{Z} . Consider the Diophantine equation

$$x^n + y^n = z^n.$$

Solving this can be notoriously difficult (PepeLaugh). However, since $\mathbb{Z}/m\mathbb{Z}$ is finite, we can use the natural ring map $\phi : \mathbb{Z} \rightarrow \mathbb{Z}/m\mathbb{Z}$ for a variety of m 's to learn about the problem.

Definition 3.6.1: Ideals

Let R be a commutative ring. An **ideal** of R is a non-empty subset $I \subseteq R$ with the following two properties:

- If $a \in I$ and $b \in I$, then $a + b \in I$.
- If $a \in I$ and $r \in R$, then $ra \in I$.

Some examples of ideals include $m\mathbb{Z} \subseteq \mathbb{Z}$ (for any $m \in \mathbb{Z}$), $\{0\} \subseteq R$ for any ring R , and R itself.

One way to create an ideal of R is to start with one element in R and take all of its multiples.

Definition 3.6.2: Principal Ideals

Let R be a commutative ring, and let $c \in R$. The **principal ideal generated by c** , denoted cR or (c) , is the set of all multiples of c ,

$$cR = (c) = \{rc \mid r \in R\}.$$

Verifying that cR is an ideal is straightforward, and left as an exercise.

The above examples still hold: $m\mathbb{Z}$, $\{0\}$, and R are all principal ideals.

In some rings (e.g. \mathbb{Z} , $\mathbb{Z}[i]$, $\mathbb{R}[x]$), every ideal is a principal ideal (such rings are called **principal ideal domains**), although this is not immediately obvious. Moreover, this is not true for rings like $\mathbb{Z}[i]$; there exist non-principal ideals.

Example 26. *Every ring has at least two ideals:*

- *the zero ideal*

$$(0) = 0R = \{0\}$$

consisting of just the zero element, and the unit ideal

$$(1) = 1R = R$$

consisting of the entire ring.

More generally, if $u \in R$ is a unit, then $uR = (u) = R$.

Just like products of rings, we can formulate a sense of building multiple ideals. A principal ideal is generated by a single element in R ; for a finite list of elements $c_1, \dots, c_n \in R$, the **ideal generated by c_1, \dots, c_n** is

$$(c_1, \dots, c_n) = c_1R + \dots + c_nR = \{r_1c_1 + \dots + r_nc_n \mid r_1, \dots, r_n \in R\}.$$

Verifying that this is an ideal is left as an exercise.

Remark 5. (*Proof Technique*) Let I be an ideal of R . If $1 \in I$, then for every $r \in R$ we have $r \cdot 1 = r \in I$, so $I = R$ is the unit ideal. Conversely, if you can construct an ideal I and prove $I = R$, we can often exploit this fact by using $1 \in I$.

Now, we can start crafting quotient rings R/I by identifying pairs of R if their difference is in I , just like we did when defining $\mathbb{Z}/m\mathbb{Z}$:

For an element $a \in R$, the set of $b \in R$ that are equivalent to a consists of the set of b such that $a - b \in I$, or equivalently, such that b is in the set $a + I$.

This parallel to the ring of integers modulo m prompts the following definitions.

Definition 3.6.3: Cosets of Ideals

Let R be a commutative ring, and let I be an ideal of R . Then for every element $a \in R$, the **coset of a** is the set

$$a + I = \{a + c \mid c \in I\}.$$

Note that a is always an element of its coset, since $0 \in I$. If $a, b \in R$ satisfy $b - a \in I$, then people often write

$$b \equiv a \pmod{I}$$

and say “ b is congruent to a modulo I .”

Given two cosets $a + I$ and $b + I$, we define their sum and product by the formulas

$$(a + I) + (b + I) = (a + b) + I, \quad (a + I) \cdot (b + I) = (a \cdot b) + I,$$

and we denote the collection of distinct cosets by R/I .

Like $\mathbb{Z}/m\mathbb{Z}$, we wish to turn R/I into a ring; it turns out that the above definitions successfully define a commutative ring.

Proposition 3.6.1

Let R be a commutative ring, and let I be an ideal of R .

1. Let $a + I, a' + I$ be two cosets. Then $a + I = a' + I$ if and only if $a' - a \in I$.
2. Addition and multiplication of cosets is well-defined, in that it doesn't matter which element of the coset we use in the definition.
3. Addition and multiplication of cosets in R/I turn R/I into a commutative ring.^a

^aTo be precise, we must require that $I \neq R$, since if $I = R$, then R/I only has one element, and we don't allow rings to have $1 = 0$.

Proof. 1. Suppose $a + I = a' + I$. Then $a' = a + c$ for some $c \in I$. But then $a' - a = c \in I$, and so $a' - a \in I$.

Now, suppose $a' - a \in I$. Then $a' - a = c \in I$ for some $c \in I$, so $a' = a + c$, $a = a' - c$. Thus any $\alpha \in a' + I$ can be written as $a' + b = a + (b + c) \in a + I$, and any $\beta \in a + I$

can be written as $\beta = a+b = a' + (b-c) \in a' + I$. Hence $a+I \subseteq a'+I$, $a'+I \subseteq a+I$, so $a+I = a'+I$.

2. For addition: let $a, b, a', b' \in R$ be elements that satisfy $a \equiv a' \pmod{I}$, $b \equiv b' \pmod{I}$; that is, $a' - a \in I$, $b' - b \in I$. By (1), $a+I = a'+I$, $b+I = b'+I$. Then

$$a' = a + s, \quad b' = b + t$$

for some $s, t \in I$. But then $a' + b' = a + s + b + t = (a + b) + (s + t)$; and since $s + t \in I$, we have $a' + b' \equiv a + b \pmod{I}$. Alternatively, $(a' + b') - (a + b) \in I$, so by (1) $(a + b) + I = (a' + b') + I$.

For multiplication: let $a, b, a', b' \in R$ be elements whose cosets satisfy

$$a' + I = a + I \text{ and } b' + I = b + I.$$

The assumption that $a + I = a' + I$ means that there is some $c \in I$ such that $a' = a + c$, and similarly the assumption that $b + I = b' + I$ means that there is some $d \in I$ such that $b' = b + d$ (since $a' \in a + I$ means that a' is of the form $a + c$ for some $c \in I$). It follows that

$$a'b' = (a + c)(b + d) = ab + \underbrace{ad + cb + cd}_{\text{This is in } I, \text{ since } c, d \in I}.$$

Since $c, d \in I$, $ad + cb + cd$ is also in I , and so $a'b' - ab = ad + cb + cd \in I$; and from (1), we see that $ab + I = a'b' + I$ are equal. □

Remark 6. One way to view the quotient ring R/I is to divide a “pie” of R into its individual “slices”, where each slice represents an element of R/I (for instance, $\mathbb{Z}/3\mathbb{Z}$ is divided into three slices: $3\mathbb{Z}$, $1+3\mathbb{Z}$, and $2+3\mathbb{Z}$). Alternatively, one can view a and b as “indistinguishable” if $a - b \in I$. As we think of $1 \sim 4$ in $\mathbb{Z}/3\mathbb{Z}$, we can abstractly partition $\mathbb{Z}/3\mathbb{Z}$ into colors; red might symbolize $3\mathbb{Z}$, green for $1+3\mathbb{Z}$, and blue for $2+3\mathbb{Z}$. Thus, $1, 4, 7$ would all be red, “the same element”.

Ideals and homomorphisms are closely related. Review the following proposition carefully; it will be used constantly.

Proposition 3.6.2

1. Let R be a commutative ring. Then the map

$$\pi : R \longrightarrow R/I, \quad a \longmapsto a + I$$

that sends an element to its coset is a surjective ring homomorphism with kernel I .

2. Let $\phi : R \rightarrow R'$ be a ring homomorphism.

- The kernel of ϕ is an ideal of R (recall that the kernel is $\ker(\phi) = \{a \in R \mid \phi(a) = 0_{R'}\}$).
- The homomorphism is injective if and only if $\ker(\phi) = (0)$.

- Let $I_\phi = \ker(\phi)$. There is a well-defined injective ring homomorphism

$$\bar{\phi} : R/I_\phi \longrightarrow R, \quad \bar{\phi}(a + I_\phi) = \phi(a).$$

Moreover, if ϕ is surjective, then $\bar{\phi}$ is an isomorphism.

Proof. To prove homomorphism:

- $\pi(0_R) = 0 + I = I = 0_{R/I}$.
- $\pi(a + b) = (a + b) + I = (a + I) + (b + I) = \pi(a) + \pi(b)$, by definition of addition of cosets of ideals.
- $\pi(a \cdot b) = (a \cdot b) + I = (a + I) \cdot (b + I) = \pi(a) \cdot \pi(b)$, by definition of multiplication of cosets of ideals.

Hence $\pi : R \rightarrow R/I$ is a ring homomorphism. Next, recall that $\ker(\pi) = \{a \in R \mid \pi(a) = 0_{R/I}\}$. But $0_{R/I} = 0 + I$, so we need $\pi(a) = I = 0 + I$. For any $a + I$, $a + I = 0 + I$ iff $a - 0 = a \in I$; thus $\ker(\pi) = I$. Surjectivity is trivial: for any $a + I \in R/I$, $a \in R$, so $\phi(a) = a + I$.

Now, let $\phi : R \rightarrow R'$ be a ring homomorphism. The kernel of ϕ is the set $\{a \in R \mid \phi(a) = 0_{R'}\}$. For any $a, b \in \ker(\phi)$, we have $\phi(a + b) = \phi(a) + \phi(b) = 0 + 0$, and so $a + b \in \ker(\phi)$. Next, let $c \in R$. For any $c \cdot a$, we have $\phi(c \cdot a) = \phi(c) \cdot \phi(a) = \phi(c) \cdot 0 = 0$, and so $c \cdot a \in \ker(\phi)$. Thus $\ker(\phi)$ forms an ideal of R .

To prove injectivity iff $\ker(\phi) = (0)$:

- Suppose ϕ is injective. Then for any $a \in \ker(\phi)$, $\phi(a) = 0 = \phi(0)$, so $a = 0$.
- Conversely, suppose $\ker(\phi) = (0)$. Suppose $\phi(a) = \phi(a')$. Then $\phi(a) - \phi(a') = \phi(a - a')$, so $a - a' \in \ker(\phi)$; but $\ker(\phi) = (0)$, so $a - a' = 0 \implies a = a'$. Hence ϕ is injective.

We finally prove well-definedness of $\bar{\phi}$. Suppose that $a + I_\phi = a' + I_\phi$ are two ways of writing a coset. We need to show $\bar{\phi}(a) = \bar{\phi}(a')$. The assumption of $a + I_\phi = a' + I_\phi$ allows us to write $a' = a + b$ for some $b \in I_\phi$. Thus

$$\bar{\phi}(a' + I) = \phi(a') = \phi(a + b) = \phi(a) + \phi(b) = \phi(a) + 0 = \phi(a),$$

and so $\bar{\phi}$ is well-defined. $\bar{\phi}$ being a ring homomorphism follows directly from ϕ being a ring homomorphism. To show injectivity,

$$\bar{\phi}(a + I_\phi) = \bar{\phi}(b + I) \iff \phi(a) = \phi(b) \iff \phi(a - b) = 0 \iff a - b \in I_\phi \iff a + I_\phi = b + I_\phi,$$

as desired (the second last step is accomplished because $\phi(a - b) = 0$ implies $a - b \in \ker(\phi) = I_\phi$).

Finally, if ϕ is surjective, then so is $\bar{\phi}$: ϕ surjective implies any $a' \in R'$ can be written as $\phi(a)$; but then any $\phi(a)$ can be written as $\bar{\phi}(a + I_\phi)$; thus $\phi(a) = \bar{\phi}(\pi(a)) = \bar{\phi}(a + I_\phi)$, and so $\bar{\phi}$ is surjective. We don't have to worry about π "eating up" any $a \in R$: if $a \equiv b \pmod{I_\phi}$, then $a = b + c$ for some $c \in I_\phi$, so

$$\phi(a) = \phi(b + c) = \phi(b) + \phi(c) = \phi(b) + 0 = \phi(b),$$

and so $\phi(a) = \phi(b)$. Hence any $\pi(a) = \pi(b)$ won't affect the surjectivity of $\bar{\phi}$. Therefore $\bar{\phi}$ is an isomorphism. \square

Definition 3.6.4: Characteristics of Rings

Let R be a ring, and let

$$\phi : \mathbb{Z} \longrightarrow R$$

be the unique homomorphism determined by $\phi(1) = 1_R$ (see Example 21). The kernel of ϕ is an ideal of \mathbb{Z} , and recall that every ideal of \mathbb{Z} is principal (therefore every ideal I of \mathbb{Z} is of the form $m\mathbb{Z}$); there is thus a unique integer $m \geq 0$ such that

$$\ker(\phi) = m\mathbb{Z}.$$

That integer m is called the **characteristic of the ring R** .

Another way to describe m is that m is the smallest integer such that

$$\phi(m) = \underbrace{1_R + \dots + 1_R}_{m \text{ terms}} = 0_R;$$

or, if no such m exists, then $m = 0$ (the characteristic of R is 0).

Example 27. The ring $\mathbb{Z}/m\mathbb{Z}$ has characteristic m , while the rings $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ all have characteristic 0. It follows from Proposition 3.31 that if a ring has characteristic m , then there is an injective ring homomorphism

$$\mathbb{Z}/m\mathbb{Z} \hookrightarrow R$$

(this follows since characteristic m implies $\ker(\phi) = m\mathbb{Z}$ for some $\phi : \mathbb{Z} \rightarrow R$, and every map $\mathbb{Z}/I_\phi = \mathbb{Z}/m\mathbb{Z} \rightarrow R$ is injective).

Note that when $m = 0$, we aren't "dividing by 0." In general, the quotient of a ring R by the zero ideal (0) is just R : $R/(0) = R$, since $R/(0)$ means we identify elements $a, b \in R$ as congruent if $a - b = 0$; or, $a = b$. Thus R has infinitely many cosets $a + (0)$, and so $R/(0) = R$.

Now, we move on to the Freshman's Dream: Many freshmen hope for $(a+b)^n = a^n + b^n$. Clearly, this isn't true in \mathbb{R} ; but is there a ring in which this is true? It turns out that yes, any ring with prime characteristic possesses that possibility!

Theorem 3.6.1: Freshman's Dream

Let p be a prime, and let R be a commutative ring with characteristic p . Then the map

$$f : R \longrightarrow R, \quad f(a) = a^p$$

is a ring homomorphism, called the **Frobenius homomorphism of R** . In particular, for all $a, b \in R$ and $n \geq 0$, we have

$$(a + b)^{p^n} = a^{p^n} + b^{p^n}.$$

Proof. Clearly $f(1) = 1$, and preserving multiplication is clear:

$$f(ab) = (ab)^p = (a^p)(b^p) = f(a)f(b).$$

For addition, we use the Binomial Theorem:

$$\begin{aligned} f(a+b) &= (a+b)^p \\ &= \sum_{k=0}^p \binom{p}{k} a^{p-k} b^k \\ &= a^p + b^p + \sum_{k=1}^{p-1} \binom{p}{k} a^{p-k} b^k. \end{aligned}$$

For any $1 \leq k \leq p-1$, $\binom{p}{k}$ is a multiple of p :

$$\binom{p}{k} = \frac{p!}{k!(p-k)!} = \frac{(p-k+1) \cdots (p-1)(p)}{k!}.$$

Moreover, since p is prime and $1 \leq k \leq p-1$, none of $k!$ will cancel with p (see Proposition 1.47 in the book). Since we're working in a ring with characteristic p , that means all the $\binom{p}{k}$ will be 0. Hence the entire sum disappears, and we are left with

$$f(a+b) = a^p + b^p = f(a) + f(b).$$

Thus f is a ring homomorphism.

To prove the freshman's dream, we use induction: the case $n = 1$ is already proven, so suppose $n = k$ is true. Then

$$(a+b)^{p^{k+1}} = ((a+b)^{p^k})^p = (a^{p^k} + b^{p^k})^p = (a^{p^k})^p + (b^{p^k})^p = a^{p^{k+1}} + b^{p^{k+1}}.$$

Thus the freshman dream holds for any $(a+b)^{p^n}$. \square

§3.7 Prime Ideals and Maximal Ideals

Primes hold great significance in number theory. Recall that an integer p is prime if and only if its only (positive) divisors are 1 and p . Moreover, if p divides any ab , then either p divides a or p divides b . We can rephrase this divisibility property for any arbitrary ideal: if a product ab is in the ideal $p\mathbb{Z}$, then either $a \in p\mathbb{Z}$ or $b \in p\mathbb{Z}$. Additionally, we note that integral domains and fields are of particular interest; thus, we wish to find for which ideals I is the quotient ring R/I an integral domain, and for which is it a field.

Definition 3.7.1: Prime Ideals

Let R be a commutative ring. An ideal I of R is a **prime ideal** if $I \neq R$ and if, whenever $ab \in I$, then either $a \in I$ or $b \in I$.

The contrapositive is also important: if I is a prime ideal, and both $a \notin I$ and $b \notin I$, then we have $ab \notin I$.

Example 28. Let $m \neq 0$. The ideal $m\mathbb{Z}$ is a prime ideal if and only if $|m|$ is a prime number.

Example 29. Let F be a field. For every $a, b \in F$ with $a \neq 0$, the principal ideal $(a + bx)F[x]$ is a prime ideal. For every $a, b, c \in F$ with $a \neq 0$ and $b^2 - ac \neq d^2$ for any $d \in F$ ($b^2 - ac$ is not the square of any element), the principal ideal $(a^2 + bx + c)F[x]$ is a prime ideal.

The largest possible ideal of R that is not R itself is also of interest.

Definition 3.7.2: Maximal Ideals

Let R be a commutative ring. An ideal I is a **maximal ideal** if $I \neq R$ and there is no ideal “larger” than I , or “contained” between I and R ; that is, for any ideal J , if $I \subseteq J \subseteq R$, then either $J = I$ or $J = R$.

Example 30. Let $p \in \mathbb{Z}$ be a prime number. Then the ideal $p\mathbb{Z}$ is both a prime ideal and a maximal ideal. This follows by combining Proposition 3.5.2 (3.17 in the textbook), which says that $\mathbb{Z}/p\mathbb{Z}$ is a field, and the following Theorem 3.7.1 (3.40 in the textbook), which says that in general, R/I is a field if and only if I is a maximal ideal.

Example 31. In the ring $\mathbb{Z}[x]$ of polynomials with integer coefficients, the principal ideals $2\mathbb{Z}[x]$ and $x\mathbb{Z}[x]$ are prime ideals, but not maximal ideals, since they are properly contained in the ideal

$$\{2a(x) + xb(x) \mid a(x), b(x) \in \mathbb{Z}[x]\}$$

generated by 2 and x , and one can check that this ideal is not all of $\mathbb{Z}[x]$. Indeed, it is a non-principal maximal ideal.

Just as prime numbers in \mathbb{Z} form the basic building blocks for all numbers (recall that any integer can be represented uniquely as a product of prime numbers), the prime and maximal ideals of R are, in some sense, the basic building blocks underlying the algebraic (and geometric) structure of R .

On the other hand, integral domains and fields are two particularly nice examples of rings. Thus, the next theorem is both interesting and important.

Theorem 3.7.1

Let R be a commutative ring, and let I be an ideal with $I \neq R$.

1. I is a prime ideal if and only if the quotient ring R/I is an integral domain.
2. I is a maximal ideal if and only if the quotient ring R/I is a field.

Proof. We first deal with the first statement:

- Suppose I is a prime ideal. Let $a+I$ and $b+I$ be elements of R/I . If $(a+I)(b+I) = 0+I = I$, then either $a+I = I$ or $b+I = I$. Thus at least one of $a+I$ and $b+I$ is equal to $0+I$, and so R/I is an integral domain (since for any $a, b \in R/I$, if $a \cdot b = (a+I)(b+I) = 0_{R/I} = 0+I$, then one of $a+I$, $b+I$ is $0_{R/I}$).
- Conversely, suppose R/I is an integral domain. Then for any $a+I, b+I \in R/I$, if $(a+I)(b+I) = 0+I = I$, then either $a+I = I$ or $b+I = I$. This implies either $a \in I$ or $b \in I$ (since $a+I = 0+I$ implies $a-0 = a \in I$).

Now, we move on to the second:

- Suppose I is a maximal ideal. Let $a+I$ be a non-zero element of R/I (e.g. $a+I \neq I$). To exploit the fact that I is a maximal ideal, construct an ideal J where

$$J = \{ar + b \mid r \in R, b \in I\}$$

(J being an ideal is left as an exercise). Taking elements of J with $r = 0$ shows that $I \subset J$, while taking $r = 1, b = 0$ shows that $a \in J$. Since $a \notin I$, J is strictly larger than I ; and since I is a maximal ideal and $I \subsetneq J \subseteq R$, we have that $J = R$. In particular, $1 \in J$. Thus there exists some $c \in R, b \in I$ such that $1 = ac + b$. In terms of R/I , and using the fact that $b \in I$ implies $b+I = I$, we have

$$1 + I = (ac + b) + I = (ac + 0) + I = ac + I = (a + I)(c + I).$$

Hence $a+I$ has a multiplicative inverse in R/I , and by definition R is commutative, so R/I is a field.

- Conversely, suppose R/I is a field. Let J be an ideal such that $I \subseteq J \subseteq R$. If $J = I$, we are done, so suppose $J \neq I$. Then for some $a \in J$, we have $a \notin I$. Then the coset $a+I \neq I = 0+I$, so $a+I$ is a non-zero element of R/I . Since R/I is a field, $a+I$ has a multiplicative inverse $c+I$ for some $c \in R$; that is,

$$1 + I = (a + I) \cdot (c + I) = ac + I,$$

so there is an element $b \in I$ such that $1 = ac + b$. But J is an ideal, so $a \in J$ implies $ac \in J$. Additionally, since $b \in I \subset J$, we have $ac + b = 1 \in J$. But since $1 \in J$, any $r \in R$ is in J : $r \cdot 1 = r \in J$. Hence $J = R$, and so I is a maximal ideal. □

This theorem leads to a very slick corollary:

Corollary 3.7.1

Every maximal ideal is a prime ideal.

Proof. If an ideal I is a maximal ideal, then R/I is a field. But R/I is a field implies that R/I is an integral domain, and thus I is a prime ideal as well. □

Note that the converse does not always hold; Example 31 (3.39 in the textbook) illustrates prime non-maximal ideals.

Remark 7. *It would be nice to say that any ring has at least one maximal ideal. It turns out that this is yet another statement equivalent to the Axiom of Choice!*

Chapter 4

Vector Spaces: Part 1

§4.1 Introduction to Vector Spaces

We studied vectors on a real plane (e.g. 2D or 3D space) in the guise of arrows; for instance, $(1, 2)$ denotes an arrow with tail at the origin and head at the coordinate $(1, 2)$. Additionally, we could add vectors “tip-to-tail, ” by concatenating one vector’s tail to another’s head; and we could multiply them by a scalar (e.g. $-2(1, 2)$ becomes $(-2, -4)$). These operations are represented as follows:

- $(a_1, b_1) + (a_2, b_2) = (a_1 + a_2, b_1 + b_2)$
- $c(a, b) = (ca, cb)$

Seem familiar? Abstractly, vectors are two types of operations: *scalar multiplication* and *vector addition*; and we can combine them using the all-powerful distributive law:

$$c(\vec{v}_1 + \vec{v}_2) = c\vec{v}_1 + c\vec{v}_2.$$

Now, we formalize this concept into a **vector space**.

§4.2 Vector Spaces and Linear Transformations

The numbers we used before $((1, 2))$ are real numbers; but that’s not necessary. Generally, we can use any sort of “numbers” that allow us to add, subtract, multiply, and divide. As discussed before, these numbers live in **fields**.

Definition 4.2.1: Fields

A **field** F is a commutative ring with the property that every non-zero $a \in F$ has a multiplicative inverse. In other words, for every $a \in F$ with $a \neq 0$, there is a $b \in F$ such that $ab = 1$; or, $F^* = F \setminus \{0\}$.

Familiar fields include \mathbb{Q} , \mathbb{R} , and \mathbb{C} ; additionally, for every prime p , there is a field $\mathbb{Z}/p\mathbb{Z}$ (see Proposition 3.17).

For vector spaces, we fix an underlying field F , and use it as a basic building block for vector spaces. Subsequently, we will use vector spaces to deeply study fields and field extensions.

Definition 4.2.2: Vector Spaces

Let F be a field. A **vector space with a field F** , denoted V_F , or alternatively a **F -vector space**, is an Abelian group V (with group law “vector addition”), together with a rule for multiplying a vector $\vec{v} \in V$ by a scalar $c \in F$ to obtain a new vector $c\vec{v} \in V$. Vector addition and scalar multiplication are required to

satisfy these axioms (in addition to the Abelian group axioms under addition and multiplication):

1. **Identity Law:** $1\vec{v} = \vec{v}$ for all $\vec{v} \in V$.
2. **Distributive Law 1:** $c(\vec{v}_1 + \vec{v}_2) = c\vec{v}_1 + c\vec{v}_2$ for all $c \in F$, $\vec{v}_1, \vec{v}_2 \in V$.
3. **Distributive Law 2:** $(c_1 + c_2)\vec{v} = c_1\vec{v} + c_2\vec{v}$ for all $c_1, c_2 \in F$, $\vec{v} \in V$.
4. **Associative Law:** $(c_1c_2)\vec{v} = c_1(c_2\vec{v})$ for all $c_1, c_2 \in F$, $\vec{v} \in V$.

The zero element of V is the **zero vector**, denoted $\mathbf{0}$. It is **not** $0 \in F$, the zero element of F .

Just as with the axiomatic definition of groups and rings, some basic facts about vector spaces can be proven directly from their definitions.

Proposition 4.2.1

Let V_F be a vector space.

- $0v = \mathbf{0}$ for all $v \in V$.
- $(-1)v = -v$ for all $v \in V$.
- Every vector $v \in V$ has a unique inverse $-v$.

Proof. • $0v = (0 + 0)v = 0v + 0v$; adding $-0v$ to both sides yields $\mathbf{0} = 0v$.

- $v + (-1)v = 1v + (-1)v = (1 + -1)v = 0v = \mathbf{0}$. Hence $(-1)v = -v$.
- Suppose $v_1, v_2 \in V$ are two inverses of $v \in V$. Then

$$v_1 = v_1 + \mathbf{0} = v_1 + (v + v_2) = (v_1 + v) + v_2 = \mathbf{0} + v_2 = v_2.$$

Hence $v_1 = v_2$. □

Vector spaces are characterized by vector addition and scalar multiplication, so we are interested in maps that preserve these properties:

Definition 4.2.3: Linear Transformations

Let F be a field, and let V, W be F -vector spaces. A **linear transformation** L from V to W is a function

$$L : V \longrightarrow W$$

satisfying

$$L(c_1\vec{v}_1 + c_2\vec{v}_2) = c_1L(\vec{v}_1) + c_2L(\vec{v}_2)$$

for all $c_1, c_2 \in F$, $\vec{v}_1, \vec{v}_2 \in V$. To specify a field, we call L an **F -linear transformation**.

Remark 8. *An alternative, and perhaps better, name for a linear transformation would be a **vector space homomorphism**, given its similarities to group homomorphisms and ring homomorphisms. However, due to historical reasons linear transformations has stuck.*

§4.3 Interesting Examples of Vector Spaces

We've already seen vectors in \mathbb{R}^2 and \mathbb{R}^3 (our world!... maybe?). More generally, we can form a vector space using a set of n -tuples with coordinates in any field.

Example 32. *Let F be a field, with $n \geq 1$ an integer. Then F^n is the F -vector space whose vectors are n -tuples of elements of F ,*

$$F^n = \{(a_1, a_2, \dots, a_n) \mid a_i \in F\}.$$

Vector addition and scalar multiplication are done coordinate-wise. Among examples of F^n vector spaces include \mathbb{R}^n , \mathbb{C}^n , and \mathbb{F}_p^n . Notice that \mathbb{F}_p^n is finite; it has exactly p^n different vectors.

Example 33. *The maps*

$$\begin{aligned} L(a_1, a_2) &= (3a_1 - 5a_2, 2a_1 + 3a_2) \\ L'(a_1, a_2, a_3) &= (3a_1 - 5a_2 + 2a_3, 2a_1 + 3a_2 - 7a_3) \end{aligned}$$

are examples of linear transformations (as one should verify!).

Example 34. *The set of polynomials $F[x]$ over a field F is a vector space, with polynomial addition and scalar multiplication the usual way. For any $a \in F$, the **evaluation map***

$$E_a : F[x] \longrightarrow F, \quad E_a(f(x)) = f(a),$$

is a linear transformation. More generally, for any list of values $a_1, \dots, a_n \in F$, we can define a linear transformation

$$E_a : F[x] \longrightarrow F^n, \quad E_a(f(x)) = (f(a_1), \dots, f(a_n)).$$

One should verify that this is a linear transformation.

Example 35. *Linear algebra can be connected with calculus! Let*

$$V = \{\text{functions } f : \mathbb{R} \rightarrow \mathbb{R}\}$$

$$V^{\text{cont}} = \{\text{continuous functions } f : \mathbb{R} \rightarrow \mathbb{R}\}$$

$$V^{\text{diff}} = \{\text{differentiable functions } f : \mathbb{R} \rightarrow \mathbb{R}\}.$$

These are \mathbb{R} -vector spaces, with function addition and scalar multiplication as usual:

$$(f + g)(x) = f(x) + g(x), \quad (cf)(x) = cf(x).$$

Differentiation is a linear transformation

$$D : V^{\text{diff}} \longrightarrow V, \quad D(f(x)) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}.$$

Similarly, for any $a \in \mathbb{R}$, integration is a linear transformation

$$I_a : V^{\text{cont}} \longrightarrow V^{\text{diff}}, \quad I_a(f(x)) = \int_a^x f(t)dt.$$

The Fundamental Theorem of Calculus can (almost; notice domain discrepancy) then be summarized by the two formulas

$$I_a \circ D(f(x)) = f(x) - f(a) \quad \text{and} \quad D \circ I_a(f(x)) = f(x).$$

§4.4 Bases and Dimension

Conveniently, every vector in \mathbb{R}^2 can be uniquely expressed using the two vectors $e_1 = (1, 0)$, $e_2 = (0, 1)$. More precisely, any vector $(a, b) \in \mathbb{R}^2$ can be written as

$$(a, b) = ae_1 + be_2,$$

and the coefficients a, b uniquely identify the vector v . A similar construction works for \mathbb{R}^n (and indeed F^n). We now axiomatize this notion to any vector space.

Definition 4.4.1: Bases

Let V_F be a vector space. A **finite^a basis for V** is a finite set of vectors $\mathcal{B} = \{v_1, \dots, v_n\} \subseteq V$ with the following property:

Every vector $\vec{v} \in V$ can be **uniquely** written in the form

$$\vec{v} = a_1v_1 + \dots + a_nv_n$$

for $a_1, \dots, a_n \in F$.

An expression of the form $a_1v_1 + \dots + a_nv_n$ is called a **linear combination** of v_1, \dots, v_n .

^anot every basis need be finite!

Example 36. Let F be a field. The **standard basis for F^n** is the collection of vectors $\{e_1, \dots, e_n\}$ where

$$e_k = (0, \dots, 1, \dots, 0)$$

the k -th coordinate is 1, and all others 0. The vector $\vec{v} = (a_1, \dots, a_n) \in F^n$ can be written uniquely by the sum

$$\vec{v} = a_1 e_1 + \dots + a_n e_n.$$

The coefficients a_1, \dots, a_n are the **coordinates of \vec{v} for the standard basis of F^n** .

The uniqueness property is important. For instance, while every vector in \mathbb{R}^2 can be expressed by some linear combination of $(1, 0)$, $(0, 1)$, and $(1, 1)$, the representation is not unique. How can we tell if a given set of vectors is a basis? Two important concepts must thus be defined, that can be used to readily check for basis properties.

Definition 4.4.2: Span and Linear Independence

Let V_F be a vector space, and let $\mathcal{A} = \{\vec{v}_1, \dots, \vec{v}_n\}$ be a set of vectors in V .

- The set \mathcal{A} **spans** V if every vector in V is a linear combination of vectors in \mathcal{A} ; that is, for any vector $\vec{v} \in V$, there exist scalars a_1, \dots, a_n such that

$$\vec{v} = a_1 v_1 + \dots + a_n v_n.$$

In general, we denote the **span of \mathcal{A}** by

$$\text{span}(\mathcal{A}) = \{a_1 v_1 + \dots + a_n v_n \mid a_1, \dots, a_n \in F\}.$$

If \mathcal{A} spans V , we say $\text{span}(\mathcal{A}) = V$.

- The set \mathcal{A} is **linearly independent** if the only scalars $a_1, \dots, a_n \in F$ that satisfy

$$a_1 v_1 + \dots + a_n v_n = \mathbf{0}$$

are $a_1 = \dots = a_n = 0$. A set is **linearly dependent** if it is not linearly independent.

Proposition 4.4.1: Basis Conditions

Let V_F be a vector space, and let $\mathcal{A} = \{v_1, \dots, v_n\}$ be a set of vectors in V . Then \mathcal{A} is a basis for V if and only if \mathcal{A} spans V and is linearly independent.

Proof. Suppose \mathcal{A} is a basis. By definition, \mathcal{A} spans V , and since

$$a_1 v_1 + \dots + a_n v_n = 0v_1 + \dots + 0v_n = \mathbf{0},$$

uniqueness of coefficients necessarily forces $a_1 = \dots = a_n = 0$. Hence \mathcal{A} is linearly independent.

Conversely, suppose \mathcal{A} spans V and is linearly independent. \mathcal{A} spanning V means every vector $v \in V$ can be expressed by a linear combination of vectors in \mathcal{A} , so it remains

to show uniqueness. Suppose $a_1v_1 + \dots + a_nv_n = b_1v_1 + \dots + b_nv_n = \mathbf{0}$. Then linear independence tells us that each $a_i - b_i$ is

$$\mathbf{0} = v - v = (a_1 - b_1)v_1 + \dots + (a_n - b_n)v_n$$

is 0; that is, $a_i - b_i = 0$, so $a_i = b_i$. Hence uniqueness is satisfied, and so \mathcal{A} is a basis for V . \square

A basis for a vector space provides the building block for the entire space, so we would like to know if a basis exists. The following theorem lets us find a basis within a spanning set.

Theorem 4.4.1: Goldilock's Theorem

Let V_F be a vector space, let \mathcal{S} be a finite set of vectors in V , and let $\mathcal{L} \subseteq \mathcal{S}$ be a linearly independent subset of \mathcal{S} (\mathcal{L} can be the empty set.) Then there is a basis \mathcal{B} for V satisfying

$$\mathcal{L} \subseteq \mathcal{B} \subseteq \mathcal{S}.$$

Rephrasing, every spanning set of V contains a basis of V , and every linearly independent set in V can be extended to a basis. We call it Goldilocks Theorem, since although \mathcal{S} may span V , it could be too big, and although \mathcal{L} may be linearly independent, it could be too small. Thus, Goldilocks builds a basis \mathcal{B} that's "just right", i.e. simultaneously "big enough" to span V and "small enough" to be linearly independent.

Proof. We look at the collections of subsets in \mathcal{S} that contain \mathcal{L} and are linearly independent:

$$\{\mathcal{A} \mid \mathcal{L} \subseteq \mathcal{A} \subseteq \mathcal{S}, \mathcal{A} \text{ is linearly independent}\}.$$

This collection is non-empty, since it always has \mathcal{L} . Choose \mathcal{B} = the subset \mathcal{A} with the largest number of elements. \mathcal{B} has the following properties:

- \mathcal{B} is a linearly independent subset of \mathcal{S} that contains \mathcal{L} .
- There are no linearly independent subsets of \mathcal{S} that contain \mathcal{L} and have more elements than \mathcal{B} .

We claim \mathcal{B} is a basis. By construction, \mathcal{B} is linearly independent, so it remains to show spanning.

- First, we show $\mathcal{S} \subset \text{span}(\mathcal{B})$. Take an arbitrary element $v \in \mathcal{S}$. If $v \in \mathcal{B}$, then v is certainly in the span of \mathcal{B} . Otherwise, $\mathcal{B} \cup \{v\}$ is strictly larger than \mathcal{B} , so by the second property $\mathcal{B} \cup \{v\}$ cannot be linearly independent. Thus, with $\mathcal{B} = \{v_1, \dots, v_m\}$, there exists non-zero $a_1, \dots, a_m, b \in F$ such that

$$a_1v_1 + \dots + a_mv_m + bv = 0.$$

$b \neq 0$, since v_1, \dots, v_m is linearly independent; so

$$v = -\frac{a_1}{b}v_1 - \dots - \frac{a_m}{b}v_m \in \text{span}(\mathcal{B}).$$

Hence $\mathcal{S} \subset \text{span}(\mathcal{B})$.

- Next, we show that $\text{span}(\mathcal{S}) \subseteq \text{span}(\mathcal{B})$. This follows from the more general fact that if $\mathcal{A}_1, \mathcal{A}_2$ are any two finite subsets of V , then

$$\mathcal{A}_1 \subset \text{span}(\mathcal{A}_2) \implies \text{span}(\mathcal{A}_1) \subseteq \text{span}(\mathcal{A}_2).$$

This is true because if any vector $v \in \mathcal{A}_1$ is in the span of \mathcal{A}_2 , then we can rewrite any linear combination $v = a_1 v_1 + \dots + a_m v_m$ of vectors in \mathcal{A}_1 in terms of vectors of \mathcal{A}_2 (since each individual v_i is in $\text{span}(\mathcal{A}_2)$); hence $\text{span}(\mathcal{A}_1) \subseteq \text{span}(\mathcal{A}_2)$.

This thus shows that $\text{span}(\mathcal{S}) \subseteq \text{span}(\mathcal{B})$; but $\text{span}(\mathcal{S}) = V$, so \mathcal{B} spans V . Since we've shown that \mathcal{B} spans V and is linearly independent (by construction), we have, by Proposition 4.4.1, that \mathcal{B} is a basis for V .

□

Remark 9. *Does every vector space admit a basis? We must first understand what it means for an infinite set \mathcal{B} to be a basis. We say such a set is a basis if every vector $v \in V$ can be written uniquely as a linear combination of some finite subset of \mathcal{B} . We require a finite subset, since in general there's no way to compute infinite sums. With this definition, the assertion that every vector space has a basis is another statement equivalent to the Axiom of Choice.*

We now move on to dimension, and a result of great importance in applying vector spaces to other areas of mathematics.

Definition 4.4.3: Dimension

Let V be a vector space with a finite basis. The **dimension** of V , denoted $\dim(V)$, is the number of vectors in a basis of V . A vector space without a finite basis is called **infinite-dimensional**.

Example 37. *The dimension of F^n is n ; the standard basis has n elements. For any $n \geq 0$, the set of polynomials*

$$\{f(x) \in F[x] \mid \deg(f) \leq n\}$$

is an F -vector space with dimension $n + 1$. The set $\{1, x, \dots, x^n\}$ is a basis.

We start with a lemma that builds up to an important result:

Lemma 4.4.1: Swap Lemma

Let V_F be a vector space, let \mathcal{S} be a spanning set of V , and let \mathcal{L} be a linearly independent set of vectors in V . Then given any vector $v \in \mathcal{L} \setminus \mathcal{S}$, we can find a vector $w \in \mathcal{S} \setminus \mathcal{L}$ such that

$$(\mathcal{S} \setminus \{w\}) \cup \{v\}$$

is still a spanning set.

In other words, we can swap a vector in \mathcal{L} but not in \mathcal{S} with a vector in \mathcal{S} but not in \mathcal{L} , while preserving the spanning set property.

Proof. We use a vector $v \in \mathcal{L} \setminus \mathcal{S}$ to build two sets of vectors:

$$(\mathcal{L} \cap \mathcal{S}) \cup \{v\} \subset \mathcal{S} \cup \{v\}.$$

The left set is linearly independent, since it is a subset of \mathcal{L} ; the right set is clearly still a spanning set. Thus, by Goldilocks Theorem, there exists a basis \mathcal{B} in between

$$(\mathcal{L} \cap \mathcal{S}) \cup \{v\} \subseteq \mathcal{B} \subseteq \mathcal{S} \cup \{v\}.$$

Further, since \mathcal{S} is a spanning set, and $v \notin \mathcal{S}$, the larger set $\mathcal{S} \cup \{v\}$ cannot be linearly independent (since v can be written as a linear combination of vectors in \mathcal{S}). Hence \mathcal{B} is not equal to $\mathcal{S} \cup \{v\}$, so there exists some $w \in \mathcal{S} \cup \{v\}$ such that $w \notin \mathcal{B}$. In particular, $w \neq v$, since $v \in \mathcal{B}$, $w \notin \mathcal{B}$. It follows that

$$\mathcal{B} \subseteq (\mathcal{S} \setminus \{w\}) \cup \{v\},$$

since $w \notin \mathcal{B}$ and $\mathcal{B} \subseteq \mathcal{S}$ implies $\mathcal{B} \subseteq \mathcal{S} \setminus \{w\}$ also. In other words, the set $(\mathcal{S} \setminus \{w\}) \cup \{v\}$ contains a basis, so it certainly is a spanning set. \square

We can then use the swap lemma to show that no linearly independent set is larger than a spanning set.

Lemma 4.4.2

Let V_F be a vector space, let $\mathcal{S} \subset V$ be a finite spanning set of V , and let $\mathcal{L} \subset V$ be a linearly independent set. Then

$$|\mathcal{L}| \leq |\mathcal{S}|.$$

Proof. If $\mathcal{L} \subseteq \mathcal{S}$, this is clearly true. Otherwise, we can find a vector $v \in \mathcal{L} \setminus \mathcal{S}$, a vector in \mathcal{L} but not \mathcal{S} . The Swap Lemma says we can find a $w \in \mathcal{S} \setminus \mathcal{L}$, a vector in \mathcal{S} but not \mathcal{L} , so that $(\mathcal{S} \setminus \{w\}) \cup \{v\}$ is still a spanning set. Let

$$\mathcal{S}' = (\mathcal{S} \setminus \{w\}) \cup \{v\};$$

note that $|\mathcal{S}'| = |\mathcal{S}|$. Also, note that \mathcal{S}' and \mathcal{L} now share an extra vector in common; or,

$$|(\mathcal{S}' \cap \mathcal{L})| \geq 1 + |(\mathcal{S} \cap \mathcal{L})|.$$

If $\mathcal{L} \subseteq \mathcal{S}'$, we are done; otherwise, we repeat the swapping process, swapping an element in \mathcal{L} but not in \mathcal{S}' into \mathcal{S}' , and removing an element from \mathcal{S}' . With each repetition, the number of shared elements between \mathcal{L} and \mathcal{S}'' increase, until eventually $\mathcal{L} \subseteq \tilde{\mathcal{S}}$ and $|\tilde{\mathcal{S}}| = |\mathcal{S}|$. Moreover, we can't repeat the process forever, since there are only finitely many elements in \mathcal{L} that can be swapped (otherwise, we would have a spanning set $\tilde{\mathcal{S}}$ that spans V and contains only elements of \mathcal{L} ; yet we would have another $v \in \mathcal{L}$ to swap. This would imply that $\tilde{\mathcal{S}} \cup \{v\}$ is linearly independent; but that's not possible, since $\tilde{\mathcal{S}}$ already spans V using only elements in \mathcal{L}).

Thus $|\mathcal{L}| \leq |\mathcal{S}|$. \square

Now, we arrive at a fundamental theorem regarding dimension of bases:

Theorem 4.4.2: Invariance of Dimension

Let V be a vector space with a finite basis. Then every basis for V has the same number of elements.

Proof. We note that V has at least one finite basis, \mathcal{B} . Let \mathcal{B}' be any other basis. Since they are bases, we know that \mathcal{B} spans and \mathcal{B}' is linearly independent; thus

$$|\mathcal{B}'| \leq |\mathcal{B}|.$$

Moreover, we also know that \mathcal{B} is linearly independent and \mathcal{B}' spans; thus

$$|\mathcal{B}| \leq |\mathcal{B}'|.$$

Hence $|\mathcal{B}| = |\mathcal{B}'|$.

□

Chapter 5

Fields: Part I

§5.1 Introduction to Fields

Recall that in Chapter 3, we introduced fields as commutative rings with an additional property: every non-zero element had a multiplicative inverse. We extended this in Chapter 4 by building vector spaces with fields.

Definition 5.1.1: Fields

A **field** is a commutative ring F with the property that for every $a \in F$ with $a \neq 0$, there is a $b \in F$ such that $ab = 1$.

Now, we begin studying fields in their own right. First, we look at how fields fit into each other, how to construct new fields from old fields, and describe all fields with finitely many elements. Finite fields are not just a mathematical curiosity; they play important roles in pure and applied mathematics, as well as engineering; some applications include signal processing, error correcting codes, and cryptography.

Remark 10. *Field theory was originally developed to aid the study of polynomials, such as finding the roots of a certain polynomial. Interestingly, field theory and finite group theory both originated from the study of polynomials! For more history, Google Cardano's Formula.*

§5.2 Abstract Fields and Homomorphisms

Recall the definition of a unit group:

Definition 5.2.1: Unit Groups

Let R be a commutative ring. The **unit group** of R is the group

$$R^* = \{a \in R \mid \text{there is some } b \in R \text{ such that } ab = 1\},$$

where the group law is ring multiplication.

Thus, a succinct way to characterize a field is that it is a commutative ring F satisfying

$$F^* = \{a \in F \mid a \neq 0\} = F \setminus \{0\}.$$

As for maps between fields, we obviously want them to preserve field properties. In particular, since fields are rings, we want our maps to *at least* be ring homomorphisms.

It turns out that it's enough to ensure that multiplicative inverses go to multiplicative inverses; moreover, somewhat surprisingly, maps between fields are always injective! Note that this statement is generally not true for rings; the ring homomorphism $\phi : \mathbb{Z} \rightarrow \mathbb{Z}/m\mathbb{Z}$ is very non-injective.

Proposition 5.2.1

Let F, K be fields, and let $\phi : F \rightarrow K$ be a ring homomorphism.

- The map ϕ is injective.
- Let $a \in F^*$. Then $\phi(a^{-1}) = \phi(a)^{-1}$.

Proof. • From Theorem 3.31(b)ii, we need to show that $\ker(\phi)$ is the zero ideal, so suppose there is a non-zero element $a \in \ker(\phi)$. $a \neq 0$ in a field F implies that $ab = 1_F$ for some $b \in F$; thus

$$1_K = \phi(1_F) = \phi(ab) = \phi(a)\phi(b) = 0 \cdot \phi(b) = 0_K,$$

a contradiction (since ring axioms require $1 \neq 0$). Thus $\ker(\phi)$ is the zero ideal, and so ϕ is injective.

- By definition of multiplicative inverse, and since ϕ is a homomorphism, we have

$$1_K = \phi(1_F) = \phi(a \cdot a^{-1}) = \phi(a) \cdot \phi(a^{-1}).$$

Hence $\phi(a^{-1})$ is a multiplicative inverse for $\phi(a)$; i.e. $\phi(a^{-1}) = \phi(a)^{-1}$.

Alternatively, from (a) we know the map $\phi : F^* \rightarrow K^*$ is well-defined (by injectivity, since if $\phi(a) = \phi(a')$, then $a = a'$), and by the homomorphism property, it is a group homomorphism from F^* to K^* . Thus, the fact that ϕ sends multiplicative inverses to multiplicative inverses is a result of Proposition 2.20 (which states that for group homomorphisms, $\phi(a^{-1})$ is the inverse of $\phi(a)$).

□

§5.3 Interesting Examples of Fields

Example 38. *Three fields are already familiar: \mathbb{Q}, \mathbb{R} , and \mathbb{C} . Moreover, they fit into each other:*

$$\mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

Example 39. *The following subset of \mathbb{C} is a field:*

$$\mathbb{Q}(i) = \{a + bi \mid a, b \in \mathbb{Q}\}.$$

The multiplicative inverse of a non-zero element $a + bi$ is obtained by “rationalizing the denominator:”

$$\frac{1}{a + bi} = \frac{1}{a + bi} \cdot \frac{a - bi}{a - bi} = \frac{a - bi}{a^2 + b^2} = \frac{a}{a^2 + b^2} - \frac{bi}{a^2 + b^2}.$$

In similar fashion, we can use $\sqrt{2}$ to describe a subfield of \mathbb{R} :

$$\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\},$$

with a multiplicative inverse obtained again by rationalizing the denominator:

$$\frac{1}{a + b\sqrt{2}} = \frac{1}{a + b\sqrt{2}} \cdot \frac{a - b\sqrt{2}}{a - b\sqrt{2}} = \frac{a - b\sqrt{2}}{a^2 - 2b^2} = \frac{a}{a^2 - 2b^2} - \frac{b}{a^2 - 2b^2}\sqrt{2}.$$

This is well-defined, since $\sqrt{2}$ is not a rational number, so $a^2 - 2b^2 \neq 0$ for any $a, b \in \mathbb{Q}$ as long as they are non-zero as well.

What about the set

$$\{a + b\sqrt{2} + c\sqrt{3} \mid a, b, c \in \mathbb{Q}\}?$$

This is not even a ring, since one can see $\sqrt{2} \cdot \sqrt{3} = \sqrt{6}$, which is not in the set. The larger set

$$\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \{a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6} \mid a, b, c, d \in \mathbb{Q}\}$$

is a field. It's easy to see ring properties, but proving that non-zero elements have an inverse is not quite so easy. We'll see tools to help with this later.

Example 40. In general, the ring $\mathbb{Z}/m\mathbb{Z}$ need not be a field. For example, $\mathbb{Z}/6\mathbb{Z}$ is not a field, since 2 does not have a multiplicative inverse. However, we've seen before that any ring $\mathbb{Z}/p\mathbb{Z}$, where p is a prime, is a field; we denote this \mathbb{F}_p . \mathbb{F}_p is an example of a **finite field**; it turns out that there are other finite fields. Indeed, for every prime power p^k , there is exactly one (up to isomorphism) finite field containing p^k elements.

Example 41. A **skew field**, also called a **division ring**, is a ring in which every non-zero element has an inverse, but is no longer required to be commutative. Wedderburn's Theorem states that every finite skew field must be commutative (e.g. is a field), but there are also many interesting non-commutative infinite fields. One such example is \mathbb{H} , the ring of quaternions.

§5.4 Subfields and Extension Fields

Definition 5.4.1: Subfields

Let K be a field. A **subfield of K** is a subset F of K that is itself a field using the addition and multiplication operations of K .

Definition 5.4.2: Extension Fields

Let F be a field. An **extension field of F** is a field K such that F is a subfield

of K . We write K/F to indicate that K is an extension field of F .^a

^aNote that K/F is just a piece of convenient notation. It doesn't mean that we're taking the quotient of K by F , despite its similarities to a quotient ring R/I .

Example 42. The field \mathbb{Q} is a subfield of \mathbb{R} , and thus \mathbb{R} is an extension field of \mathbb{Q} ; similarly with \mathbb{Q}, \mathbb{R} and \mathbb{C} .

The fields $\mathbb{Q}(i)$ and $\mathbb{Q}(\sqrt{2})$ are extension fields of \mathbb{Q} . The former is a subfield of \mathbb{C} but not \mathbb{R} , while the latter is a subfield of \mathbb{R} . Neither are subfields of each other.

The notation $\mathbb{Q}(i)$, $\mathbb{Q}(\sqrt{2})$ is a special case of the following general construction.

Proposition 5.4.1

Let L/F be an extension of fields, and let $\alpha_1, \dots, \alpha_n \in L$. Then there is a unique field K with the following properties:

1. $F \subseteq K \subseteq L$.
2. $\alpha_1, \dots, \alpha_n \in K$.
3. If K' is a field satisfying $F \subseteq K' \subseteq L$ and $\alpha_1, \dots, \alpha_n \in K'$, then $K \subseteq K'$.

The field K is denoted $F(\alpha_1, \dots, \alpha_n)$ and is called the **extension field of F generated by $\alpha_1, \dots, \alpha_n$** . Intuitively, it is the smallest subfield of L that contains both F and $\alpha_1, \dots, \alpha_n$.

Proof. Let S be the set consisting of all subfields of L that contain F and $\alpha_1, \dots, \alpha_n$. The set is not empty, since $L \in S$. Let K be the intersection of all of the fields in S . Then K is a field, since

$$\begin{aligned} \alpha, \beta \in K &\iff \alpha, \beta \in K' \text{ for every } K' \in S, \\ &\implies \alpha \pm \beta, \alpha\beta, \alpha^{-1} \in K' \text{ for every } K' \in S, \text{ since } K' \text{ is a field,} \\ &\implies \alpha \pm \beta, \alpha\beta, \alpha^{-1} \in K. \end{aligned}$$

Clearly, K contains F and α_i , since it is the intersection of fields that contain both; and if $K' \subseteq L$ contains both, then $K' \in S$, so K' is one of the fields whose intersection forms K . Hence $K \subseteq K'$. \square

Let K/F be an extension of fields. Observe that we can add elements of K , and multiply elements in K by elements in F ; thus, the field K becomes an F -vector space. Essentially, we discard most of the multiplication operation in K , and restrict it solely to multiplication by elements in F . This allows us to use tools from linear algebra to study field extensions.

Definition 5.4.3: Degree of Field Extensions

Let K/F be an extension of fields. The **degree of K over F** , denoted $[K : F]$, is the dimension of K viewed as an F -vector space,

$$[K : F] = \dim_F(K).$$

If $[K : F]$ is finite, we say K/F is a **finite extension**; otherwise, we say K/F is an **infinite extension**.

Example 43. The fields $\mathbb{Q}(i)$, $\mathbb{Q}(\sqrt{2})$ have degree 2 over \mathbb{Q} :

- $[\mathbb{Q}(i) : \mathbb{Q}] = 2$, since $\{1, i\}$ is a \mathbb{Q} -basis for $\mathbb{Q}(i)$.
- $[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$, since $\{1, \sqrt{2}\}$ is a \mathbb{Q} -basis for $\mathbb{Q}(\sqrt{2})$.

Similarly, we have $[\mathbb{C} : \mathbb{R}] = 2$, since $\{1, i\}$ is an \mathbb{R} -basis for \mathbb{C} . On the other hand, $[\mathbb{R} : \mathbb{Q}] = \infty$:

Proof. Suppose that $\{a_1, \dots, a_n\} \subset \mathbb{R}$ is a finite \mathbb{Q} -basis for \mathbb{R} . Then

$$\mathbb{R} = \{c_1 a_1 + \dots + c_n a_n \in \mathbb{Q}\}.$$

But the set on the right is countable (since its cardinality is the same as the cardinality of the set of n -tuples of rational numbers), while \mathbb{R} is uncountable. \square

The next theorem is similar to the index multiplication rule, which counted cosets in a chain of groups.

Theorem 5.4.1

Let $L/K/F$ be an extension of fields (that is, L is a field extension of K , while K is a field extension of F). Then

$$[L : F] = [L : K] \cdot [K : F],$$

in the sense that one of the following is true:

- All of the degrees $[L : F]$, $[L : K]$, $[K : F]$ are finite, and the above equation is true.
- $[L : F] = \infty$, and either $[L : K] = \infty$ or $[K : F] = \infty$.

Proof. We start with the case L/K and K/F are finite extensions. This means that we can choose bases

- $\mathcal{A} = \{a_1, \dots, a_m\}$, a basis for K as an F -vector space
- $\mathcal{B} = \{b_1, \dots, b_n\}$, a basis for L as a K -vector space

We claim that the mn products in the set

$$\mathcal{C} = \{a_i b_j \mid 1 \leq i \leq m, 1 \leq j \leq n\}$$

are distinct and that they form a basis for L as an F -vector space. Assuming this, it's easy to prove the equation:

$$[L : F] = \dim_F(L) = |\mathcal{C}| = mn = |\mathcal{A}| |\mathcal{B}| = \dim_F(K) \cdot \dim_K(L) = [K : F] \cdot [L : K].$$

We first prove that this is a linearly independent set. Suppose we have an F -linear combination of the elements of \mathcal{C} that sum to 0:

$$\sum_{i=1}^m \sum_{j=1}^n c_{ij} a_i b_j = 0, \quad c_{ij} \in F.$$

We wish to show that every c_{ij} is zero. We switch the order of the sums (due to commutativity) to get

$$\sum_{j=1}^n \left(\sum_{i=1}^m c_{ij} a_i \right) b_j = 0.$$

The inner sum is in K , since $c_{ij} \in F$, $a_i \in K$. This gives a linear combination of b_j with coefficients in K ; but $b_i \in \mathcal{B}$ is a basis for L as a K -vector space, so it is K -linearly independent. That is, every $c_{ij} a_i = 0$, so we have

$$\sum_{i=1}^m c_{ij} a_i = 0.$$

But we also know that $a_i \in \mathcal{A}$ is a basis for K as an F -vector space, and each c_{ij} is in F , so F -linear independence of a_1, \dots, a_m implies that every $c_{ij} \in F$ is 0, and so \mathcal{C} is a linearly independent set.

Spanning is left as an exercise to the reader. \square

§5.5 Polynomial Rings

We will now briefly discuss some properties of polynomials with coefficients inside a field F .

Definition 5.5.1: Degree of f

Let F be a field, and let $f(x) \in F[x]$ be a non-zero polynomial, where

$$f(x) = a_0 + a_1x + \dots + a_dx^d, a_d \neq 0;$$

in other words, d is the highest power with a non-zero coefficient. The **degree** of f is

$$\deg(f) = d.$$

By convention, we set $\deg(0) = -\infty$; further, if $a_d = 1$, then we say that f is a **monic polynomial**.

One can easily check that for any two polynomials $f_1, f_2 \in F[x]$, we have

$$\deg(f_1 f_2) = \deg(f_1) + \deg(f_2).$$

A useful property of polynomials is the following algorithm, which allows one to extend division into the ring of polynomials. It is analogous to the Division Algorithm for integers.

Proposition 5.5.1: Division-with-Remainder for Polynomials

Let F be a field, and let $f(x), g(x) \in F[x]$ be a polynomial with $g(x) \neq 0$. Then there are unique polynomials $q(x), r(x) \in F[x]$ satisfying

$$f(x) = g(x)q(x) + r(x) \text{ with } \deg(r) < \deg(g).$$

Proof. We prove existence by following the long-division with remainder algorithm for polynomials, and we use induction to generalize for any degree. So, if we assume that the degree of the polynomial $g(x)$ is fixed, say with

$$d = \deg(g) \geq 1,$$

our goal is to verify the statement P_n :

If $f(x) \in F[x]$ is a polynomial with $\deg(f) = n$, then there are unique polynomials $q(x), r(x) \in F[x]$ satisfying

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

We first prove that P_0, \dots, P_{d-1} are true. Let $f(x) \in F[x]$ with $\deg(f) < d$. Then we can take $q(x) = 0$, $r(x) = f(x)$, since

$$f(x) = 0 \cdot g(x) + f(x)$$

is clearly true, and $\deg(f) < d = \deg(g)$ by definition.

Now, let $n \geq d$, and assume P_0, \dots, P_{n-1} are true. Let $f(x) \in F[x]$ be a polynomial with $\deg(f) = n$; that is, we have

$$f(x) = ax^n + \dots \text{ and } g(x) = bx^d + \dots, \quad a, b \in F, \quad a, b \neq 0.$$

Then, if we multiply $g(x)$ by $\frac{a}{b}x^{n-d}$, we get a new polynomial

$$h(x) = f(x) - \frac{a}{b}x^{n-d}g(x).$$

Then $\deg(h) < n$, since we canceled out the ax^n term in $f(x)$. Then, we can use the inductive hypothesis to find $q(x), r(x) \in F[x]$ such that

$$h(x) = g(x)q(x) + r(x), \quad \deg(r) < \deg(g).$$

But then

$$\begin{aligned} f(x) &= h(x) + \frac{a}{b}x^{n-d}g(x) \\ &= g(x)q(x) + r(x) + \frac{a}{b}x^{n-d}g(x) \\ &= g(x)\left(q(x) + \frac{a}{b}x^{n-d}\right) + r(x). \end{aligned}$$

We have thus written $f(x)$ in the desired form:

$$f(x) = g(x) \cdot (\text{polynomial}) + (\text{polynomial whose degree is strictly smaller than the degree of } g(x)).$$

(since $q(x) + \frac{a}{b}x^{n-d} \in F[x]$). Uniqueness is left as an exercise. \square

Now, we move on to ideals of polynomial rings. Recall that an ideal is **principal** if it consists of multiples of a single element of R ; that is, for $c \in R$,

$$I = \{cr \mid r \in R\}.$$

This ideal is called the **principal ideal generated by c** , and is often denoted cR or (c) . Principal ideals are quite easy to work with, which makes the following theorem quite interesting and important.

Theorem 5.5.1

Let F be a field, and let $I \subseteq F[x]$ be an ideal in the ring $F[x]$. Then I is a principal ideal.

Proof. If I is the zero ideal ($I = (0)$), then I is clearly a principal ideal, so assume $I \neq (0)$; that is, I contains at least one non-zero polynomial. Let $g(x) \in I$ be the non-zero polynomial with the lowest degree. We claim $I = (g)$.

Take an arbitrary $f \in I$. From the Division Algorithm for polynomials, we know that

$$f(x) = g(x)q(x) + r(x)$$

for some $q(x), r(x) \in F[x]$, $\deg(r) < \deg(g)$. Thus

$$r(x) = f(x) - g(x)q(x) \in I,$$

since $f(x), g(x)q(x) \in I$. But $g(x)$ has the smallest possible degree of any non-zero polynomial in I , while we also have $\deg(r) < \deg(g)$. Thus $r(x) = 0$ necessarily, and so

$$f(x) = g(x)q(x) \in (g) = g(x)F[x].$$

Thus $I \subseteq (g)$; and the fact that $g \in I$ forces $(g) \subseteq I$ by ideal properties. Hence $I = (g)$ is a principal ideal. \square

§5.6 Building Extension Fields

Now, we start building fields using roots of polynomials. However, we sometimes don't know where these roots might live. So instead, we take a field F and a polynomial $f(x) \in F[x]$, and use them to construct a field extension K/F that “magically” contains a root of $f(x)$. A prototype for this construction is the abstract construction of the field of complex numbers from the real numbers and the polynomial $x^2 + 1$ (recall how we can find a map $\phi : \mathbb{R}[x]/(x^2 + 1)\mathbb{R}[x] \rightarrow \mathbb{C}, \phi(f(x)) = f(i)$, by taking the evaluation homomorphism E_i).

Definition 5.6.1: Irreducibility

Let F be a field. A non-constant polynomial $f(x) \in F[x]$ is **reducible** (over F) if there exist non-constant polynomials $g(x), h(x) \in F[x]$ such that $f(x) = g(x)h(x)$ (f factors into them). An **irreducible polynomial** is a non-constant polynomial that is not reducible, i.e., a polynomial with no non-trivial factorizations in $F[x]$; that is, the only factorizations are constant polynomials and multiples of itself (we will later generalize this to arbitrary rings).

Example 44. The polynomials $x^2 - 1$ and $x^2 - x - 2$ are reducible in $\mathbb{Q}[x]$:

$$x^2 - 1 = (x + 1)(x - 1) \text{ and } x^2 - x - 2 = (x - 1)(x - 2).$$

The polynomials $x^2 + 1$ and $x^2 - 2$ are irreducible in $\mathbb{Q}[x]$ (although the second one is equivalent to saying $\sqrt{2}$ is not rational). On the other hand, in the ring $\mathbb{Q}(i)[x]$, $x^2 + 1$ is reducible, since it can factor as

$$x^2 + 1 = (x + i)(x - i).$$

Reducibility depends on both the polynomial and the underlying field.

Example 45. *If we take finite fields as coefficients, then there are only finitely many polynomials of any given degree. For instance, only four quadratic polynomials exist in $\mathbb{F}_2[x]$:*

$$x^2 = x \cdot x, \quad x^2 + 1 = (x + 1)^2, \quad x^2 + x = x(x + 1), \quad x^2 + x + 1.$$

Only the last one is irreducible. Similarly, there are eight cubic polynomials in $\mathbb{F}_3[x]$; only $x^3 + x + 1$ and $x^3 + x^2 + 1$ are irreducible.

Remark 11. *Every polynomial of degree 1 is irreducible, and it's not hard to show that if $f(x)$ has degree 2 or 3, then it is irreducible if and only if it has no roots. However, this doesn't necessarily hold for higher degrees; for instance, given two irreducible polynomials of degree 2, $g(x)$, $h(x)$, then $g(x)h(x)$ is a reducible polynomial of degree 4, yet it has no roots in F .*

Theorem 5.6.1

Let F be a field, and let $f(x) \in F[x]$. The following are equivalent:

1. The polynomial $f(x)$ is irreducible.
2. The principal ideal $f(x)F[x]$, generated by $f(x)$, is a maximal ideal.
3. The quotient ring $F[x]/f(x)F[x]$ is a field.

Proof. The equivalence of the second and third statements is a result of Theorem 3.40 (I is a maximal ideal if and only if R/I is a field).

We start by proving $f(x)$ irreducible implies $f(x)F[x]$ is a maximal ideal. Suppose I is an ideal satisfying

$$f(x)F[x] \subseteq I \subseteq F[x].$$

We wish to show $I = F[x]$ or $I = f(x)F[x]$.

From the Theorem above, we know that every ideal in $F[x]$ is principal, so we can find some $g(x) \in I$ such that $g(x)F[x] = I$. The inclusion of fields thus becomes

$$f(x)F[x] \subseteq g(x)F[x] \subseteq F[x].$$

In particular, since $f(x)$ is in the principal ideal generated by $g(x)$, we have

$$f(x) = g(x)h(x), \quad h(x) \in F[x].$$

Now, we use the irreducibility of $f(x)$. Since $f(x)$ is irreducible, one of $g(x)$, $h(x)$ is a non-zero constant polynomial. This gives two cases:

- $g(x) \in F^*$, which tells us $I = g(x)F[x] = F[x]$ (since $g(x) \in F^*$ and g constant implies that its inverse is also in $F[x]$, so we can take $g(x)g^{-1}(x) = 1 \in I$. $1 \in I$ then forces $I = F[x]$.)

- $h(x) \in F^*$, which tells us $f(x)F[x] = g(x)h(x)F[x] = g(x)F[x] = I$ (since $h(x) \in F^*$ means we can find an inverse to cancel it out with.)

Thus $f(x)F[x]$ is a maximal ideal.

We then show the reverse: suppose $f(x)F[x]$ is a maximal ideal. Suppose $f(x)$ factors into $g(x)$ and $h(x)$ for some $g(x), h(x) \in F[x]$. We wish to show that either $g(x)$ or $h(x)$ is a constant polynomial. Consider the ideals

$$f(x)F[x] \subseteq g(x)F[x] \subseteq F[x].$$

$f(x)F[x]$ maximal means that either $f(x)F[x] = g(x)F[x]$ or $g(x)F[x] = F[x]$, so we consider both cases.

- Suppose $f(x)F[x] = g(x)F[x]$. Since $f(x) = g(x)h(x)$, we can replace $f(x)F[x]$ with $g(x)h(x)F[x]$. Then we have

$$\begin{aligned} g(x) \in f(x)F[x] = g(x)h(x)F[x] &\implies g(x) = g(x)h(x)s(x), \quad s(x) \in F[x] \\ &\implies h(x)s(x) = 1 \end{aligned}$$

[by properties of fields: $g \cdot f = g$]

Thus $h(x)$ (and $s(x)$) must necessarily be a constant polynomial.

- Suppose $g(x)F[x] = F[x]$. Then $1 \in g(x)F[x]$, so $g(x)t(x) = 1$ for some $t(x) \in F[x]$, and so $g(x)$ must be a constant polynomial.

In both cases, either $g(x)$ or $h(x)$ is a constant polynomial; hence the only way to factor $f(x) = g(x)h(x)$ is for one of them to be constant. Thus $f(x)$ is definitionally irreducible. \square

Now, given a field F and a polynomial $f(x) \in F[x]$, we wish to use this theorem to construct an extension field of F that contains a root of f . Later, we'll see a more general result; but even this one is spectacular: it says that if you're given any irreducible polynomial $f(x) \in F[x]$, if you're willing to make the field F slightly larger, then you can find a root of $f(x)$.

Theorem 5.6.2

Let F be a field, let $f(x) \in F[x]$ be an irreducible polynomial, let $I_f = f(x)F[x]$ be the principal (maximal) ideal generated by $f(x)$, and let $K_f = F[x]/I_f$ be the quotient ring.

1. The ring K_f is a field.
2. The polynomial $f(x)$ has a root in the field K_f .
3. The field K_f is a finite extension of the field F . Its degree is given by

$$[K_f : F] = \deg(f).$$

Proof. 1. The above theorem tells us that I_f is a maximal ideal of the ring $F[x]$, and Theorem 3.40b (maximal ideal iff quotient ring is field) tells us that $K_f = F[x]/I_f$ is a field.

2. (It almost seems like cheating!) We claim that

$$\beta = x + I_f \in K_f$$

is a root of $f(x)$. To see this, write $f(x)$ as

$$f(x) = b_0 + b_1x + \dots + b_dx^d.$$

Then

$$\begin{aligned} f(\beta) &= b_0 + b_1\beta + \dots + b_d\beta^d \\ &= b_0(1 + I_f) + b_1(x + I_f) + \dots + b_d(x + I_f)^d \\ &= (b_0 + I_f) + (b_1x + I_f) + \dots + (b_dx^d + I_f) \quad [\text{convince yourself why this is true, given the properties of cosets}] \\ &= b_0 + b_1x + \dots + b_dx^d + I_f \quad [\text{by coset addition}]. \end{aligned}$$

Thus $f(\beta) = 0 + I_f$, and so $f(\beta)$ is the zero element of the field K_f .

3. Let $d = \deg(f)$, and let

$$\beta = x + I_f = \text{the coset of } x \text{ in the quotient ring } F[x]/I_f.$$

We claim that the set

$$\mathcal{B} = \{1, \beta, \dots, \beta^{d-1}\}$$

forms an F -basis for K_f .

We first show that \mathcal{B} spans. Let $g(x) + I_f$ be an arbitrary element of K_f . We divide the polynomial $g(x)$ by $f(x)$ to get a quotient and remainder:

$$g(x) = f(x)q(x) + r(x), \quad q(x), r(x) \in F[x], \quad \deg(r) < \deg(f) = d.$$

Since $f(x) \in I_f$, the cosets $g(x) + I_f$ and $r(x) + I_f$ are the same. On the other hand, if we write $r(x)$ as

$$r(x) = a_0 + a_1x + \dots + a_{d-1}x^{d-1} \text{ with } a_0, \dots, a_{d-1} \in F,$$

then we can compute

$$\begin{aligned} g(x) + I_f &= r(x) + I_f \\ &= (a_0 + a_1x + \dots + a_{d-1}x^{d-1}) + I_f \\ &= a_0(1 + I_f) + a_1(x + I_f) + \dots + a_{d-1}(x + I_f)^{d-1} \\ &= a_0\beta + \dots + a_{d-1}\beta^{d-1}. \end{aligned}$$

Thus, any element of $K_f = F[x]/I_f$ can be written as an F -linear combination of the elements in \mathcal{B} , and so \mathcal{B} spans K_f .

Next, we show that \mathcal{B} is linearly independent. Suppose we have

$$c_0 + c_1\beta + \dots + c_{d-1}\beta^{d-1} = 0, \quad c_i \in F.$$

We wish to show that every c_i vanishes. Using $\beta = x + I_f$, we have

$$\begin{aligned} 0 + I_f &= c_0 + c_1\beta + \dots + c_{d-1}\beta^{d-1} \\ &= c_0(1 + I_f) + c_1(x + I_f) + \dots + c_{d-1}(x + I_f)^{d-1} \\ &= (c_0 + I_f) + (c_1x + I_f) + \dots + (c_{d-1}x^{d-1} + I_f) \\ I_f &= (c_0 + c_1x + \dots + c_{d-1}x^{d-1}) + I_f. \end{aligned}$$

This tells us that the polynomial $c_0 + c_1x + \dots + c_{d-1}x^{d-1}$ is in the principal ideal I_f generated by $f(x)$, so for some $g(x) \in F[x]$ we have

$$c_0 + c_1x + \dots + c_{d-1}x^{d-1} = f(x)g(x).$$

If $g(x) \neq 0$, we can take the degrees of both sides to obtain a contradiction:

$$d-1 \geq \deg(c_0 + \dots + c_{d-1}x^{d-1}) = \deg(f(x)g(x)) = \deg(f(x)) + \deg(g(x)) = d + \deg(g(x)) = d.$$

Hence $g(x) = 0$, and so the above equation becomes the following equality

$$c_0 + \dots + c_{d-1}x^{d-1} = 0$$

that is true in $F[x]$. Therefore $c_0 = c_1 = \dots = c_{d-1} = 0$, which proves linear independence. Thus \mathcal{B} is a basis for K_f/F , and so

$$[K_f : F] = |\mathcal{B}| = d = \deg(f).$$

□

§5.7 Finite Fields

Now, we'll apply all of our tools that we've developed to explore finite fields, and construct finite fields of various prime power orders.

Before starting, we recall the definition of a **characteristic** of a ring R :

Definition 5.7.1: Characteristic

The **characteristic** of a ring R is the integer $m \geq 0$ generating the kernel of the unique homomorphism

$$\phi : \mathbb{Z} \longrightarrow R.$$

Alternatively, m is the smallest integer such that

$$\underbrace{\alpha + \dots + \alpha}_{m \text{ terms}} = 0.$$

We know that ϕ is unique, since $\phi(1) = 1_R$, which determines $\phi(n)$ for all $n \geq 0$. Moreover, recall that every ideal of \mathbb{Z} is principal, and the kernel of any ring homomorphism is an ideal of the starting ring. Thus, when we say m generates the kernel, we mean

$$\ker(\phi) = m\mathbb{Z}.$$

Proposition 5.7.1

Let F be a finite field.

1. The characteristic of F is prime.
2. Let $p = \text{char}(F)$. Then the finite field \mathbb{F}_p is a subfield of F , in the sense that there is a unique injective homomorphism

$$\mathbb{F}_p \hookrightarrow F.$$

3. The number of elements of F is given by

$$|F| = p^{[F:\mathbb{F}_p]}.$$

In particular, the number of elements in a finite field is always the power of a prime.

Proof. 1. Let $m = \text{char}(F)$; by definition, $m\mathbb{Z}$ is the kernel of the unique homomorphism $\phi : \mathbb{Z} \rightarrow F$. It follows from Proposition 3.31 that ϕ induces an injective ring homomorphism

$$\bar{\phi} : \mathbb{Z}/m\mathbb{Z} \hookrightarrow F.$$

Moreover, F finite tells us that $m \neq 0$ (since eventually, elements will have to cycle back again, or else it would be infinite). We need to show that m is prime.

Suppose m factors into $m = kn$ (we put lines above integers to distinguish elements in \mathbb{Z} and $m\mathbb{Z}$). Now,

$$0 = \bar{\phi}(\bar{0}) = \bar{\phi}(\overline{m}) = \bar{\phi}(\overline{kn}) = \bar{\phi}(\overline{k}) \cdot \bar{\phi}(\overline{n}).$$

Since F is an integral domain (it has no zero divisors), we conclude that either

$$\bar{\phi}(\overline{k}) = 0 \text{ or } \bar{\phi}(\overline{n}) = 0.$$

But $\bar{\phi}$ is injective; thus either $\overline{k} = 0$ or $\overline{n} = 0$. In other words, $k + m\mathbb{Z} = 0 + m\mathbb{Z} = m\mathbb{Z}$ or $n + m\mathbb{Z} = m\mathbb{Z}$; hence $m \mid k$ or $m \mid n$, and thus the factorization $m = kn$ is a trivial factorization. This shows that m has no non-trivial factorizations, so it is prime.

2. We know from *a* that m is prime, so the injective map $\bar{\phi}$ gives this result.
3. From (b), we know that F is a field extension of \mathbb{F}_p . And since F has only finitely many elements, the dimension of F as an \mathbb{F}_p -vector space must be finite, since any basis is a subset of F . Let $d = [F : \mathbb{F}_p] = \dim_{\mathbb{F}_p}(F)$, and let

$$\mathcal{B} = \{\beta_1, \dots, \beta_d\}$$

be an \mathbb{F}_p -basis for F (i.e. it has elements in F , and forms a basis with coefficients in \mathbb{F}_p). The definition of basis tells us that every element of F looks like

$$c_1\beta_1 + \dots + c_d\beta_d \text{ with } c_1, \dots, c_d \in \mathbb{F}_p,$$

and that the different choices of c_1, \dots, c_d give distinct elements of F . In fancier terms, because every element in F is spanned by unique $c_1, \dots, c_d \in \mathbb{F}_p$, there is a bijection defined by

$$\mathbb{F}_p^d \longrightarrow F, (c_1, \dots, c_d) \mapsto c_1\beta_1 + \dots + c_d\beta_d.$$

Hence

$$|F| = (\mathbb{F}_p)^d = p^d.$$

Thus the number of elements in a finite field is always the power of a prime. □

The full proof of the next result is highly involved, and would lead us too far afield (hehe), so we prove up to degree 3, and briefly indicate a way to prove it for higher degrees. This result is by no means obvious: for example, it would not be true if we replaced the field \mathbb{F}_p with \mathbb{R} , since every irreducible polynomial in $\mathbb{R}[x]$ has degree 1 or 2.

Theorem 5.7.1

Let p be a prime, and let $d \geq 1$. Then the ring $\mathbb{F}_p[x]$ contains an irreducible polynomial of degree d .

Proof. For $d = 1$, any degree 1 polynomial will work.

Since we can always multiply a polynomial by a non-zero scalar without affecting its irreducibility, it suffices to look at **monic polynomials**; that is, polynomials whose leading coefficient is 1. We let

- $\text{Poly}_d = \{\text{monic polynomials in } \mathbb{F}_p[x] \text{ having degree } d\}$
- $\text{Irred}_d = \{\text{monic irreducible polynomials in } \mathbb{F}_p[x] \text{ having degree } d\}$
- $\text{Red}_d = \{\text{monic reducible polynomials in } \mathbb{F}_p[x] \text{ having degree } d\}$

We note that $|\text{Poly}_d| = p^d$, since a monic polynomial of degree d in $\mathbb{F}_p[x]$ has exactly d coefficients, each of which may freely be chosen in \mathbb{F}_p (and thus there are p^d different choices).

We start with $d = 2$, so we wish to compute $|\text{Irred}_2|$, the number of irreducible monic quadratic polynomials. We won't compute this directly, but instead use a lesson from combinatorics: we'll count the number of reducible monic quadratic polynomials (the polynomials we don't want) and subtract this value from the total number of monic quadratic polynomials.

So, which quadratic polynomials are irreducible? Precisely the polynomials that factor as

$$x^2 + ax + b = (x - a)(x - b) \text{ for some } a, b \in \mathbb{F}_p.$$

It may look like there are p^2 different choices for a, b , but we must not double count polynomials. There are two cases:

- $a = b$: then we get p different polynomials $(x - a)^2$ for some $a \in \mathbb{F}_p$.
- $a \neq b$: then there are $p(p-1)$ choices for the pair (a, b) , but since the order of a and b does not change the product $(x - a)(x - b)$, the number of different polynomials we get is $\binom{p}{2}$ (p choose 2, where order doesn't matter).

Thus,

$$|\text{Red}_2| = p + \binom{p}{2} = p + \frac{p(p-1)}{2} = \frac{p^2 + p}{2}.$$

Removing these reducible polynomials from the totality of all monic polynomials gives us

$$|\text{Irred}_2| = |\text{Poly}_2| - |\text{Red}_2| = p^2 - \frac{p^2 + p}{2} = \frac{p^2 - p}{2}.$$

This is positive for all $p \geq 2$, and so Irred_2 is non-empty.

Now, we look at cubic polynomials. We wish to count how many are irreducible. There's quite a few ways to factor a monic cubic polynomial; we list them here, along with how many distinct ones there are:

- $(x - a)^3$; $a \in \mathbb{F}_p$, so p distinct polynomials.
- $(x - a)^2(x - b)$; $a, b \in \mathbb{F}_p$, $a \neq b$, so $p(p-1)$ distinct polynomials. (Note that order **does** matter here, so we do not divide by 2).
- $(x - a)(x - b)(x - c)$; $a, b, c \in \mathbb{F}_p$, all distinct, so $\binom{p}{3}$

- $(x - a)(x^2 + bx + c)$; $a, b, c \in \mathbb{F}_p$, $x^2 + bx + c$ irreducible, so $p \cdot |\text{Irred}_2| = \frac{p^3 - p^2}{2}$.

Thus the total number of reducible monic cubic polynomials is

$$\text{Red}_3 = p + p(p - 1) + \frac{p(p - 1)(p - 2)}{6} + \frac{p^3 - p^2}{2} = \frac{2p^3 + p}{3}.$$

Hence

$$|\text{Irred}_3| = |\text{Poly}_3| - |\text{Red}_3| = p^3 - \frac{2p^3 + p}{3} = \frac{p^3 - p}{3}.$$

Thus Irred_3 is non-empty for all $p \geq 2$. □

Example 46. Taking $p = 2$ in the formulas

$$|\text{Irred}_2| = \frac{1}{2}(p^2 - 2) \text{ and } |\text{Irred}_3| = \frac{1}{3}(p^3 - p)$$

that we derived while proving the above theorem shows that $\mathbb{F}_2[x]$ has exactly one monic quadratic polynomials, and exactly two monic cubic polynomials. These values agree with an investigation in an earlier example.

We're going to use the above theorem to prove the first part of the following fundamental theorem, but for completeness we include both parts. A later section will provide an alternative proof for the first part, and an even later section will provide both yet another alternative proof and a proof for the second part.

Theorem 5.7.2

Let p be a prime, and let $d \geq 1$.

1. There exists a field F containing exactly p^d elements.
2. Any two fields containing p^d elements are isomorphic.

Proof. 1. The above theorem tells us that there is an irreducible polynomial $f(x) \in \mathbb{F}_p[x]$ satisfying $\deg(f) = d$. We let $K_f = \mathbb{F}_p[x]/f(x)\mathbb{F}_p[x]$ be the field described in Theorem 5.6.2 (5.27 in the textbook). That theorem tells us that $[K_f : \mathbb{F}_p] = \deg(f) = d$, and then Proposition 5.7.1 (5.28 in the textbook) says that

$$|K_f| = p^{[K_f : \mathbb{F}_p]} = p^d.$$

2. We will see... □