MATH 351: INTRODUCTION TO ABSTRACT ALGEBRA
SUMMER 2017, RUTGERS UNIVERSITY

KELLY SPENDLOVE

On Algebra

> *Algebra is the offer made by the devil to the mathematician. The devil says: I will give you this powerful machine, it will answer any question you like. All you need to do is give me your soul: give up geometry and you will have this marvelous machine.*
>
> – Sir Michael Atiyah[1]

'Abstract Algebra' is a study of *structure* and 'arithmetic systems', e.g. groups, rings, fields. Algebra grew out of arithmetic - abstract algebra will axiomatize basic concepts you've studied before (in $\mathbb{Z}, \mathbb{Z}_n$) and we will study their structure.

Here is a prototype of a structure/decomposition theorem (we will see this again at the end of Ch. 1)

Theorem 0.1 (Fundamental Theorem of Arithmetic)**.** *Let $n \in \mathbb{Z}$ with $n \neq 0, \pm 1$. Then n is a product of primes, i.e.*

$$n = p_1 p_2 \cdots p_n$$

for p_i prime and this factorization is unique up to reordering

The key to this class will be working through as many problems as possible. It is crucial to *read the book*. You should understand *every proof in the book*.

For the most part these notes will follow *Abstract Algebra, An Introduction* [4], T. Hungerford (3rd edition), colloquially known as 'Baby Hungerford'. ('Hungerford' typically refers to his book *Algebra* in the Graduate Text in Mathematics series).

Baby Hungerford (B.H.) is sometimes considered 'wordy', but all of those words assemble into a book that is cogent. These notes will likely not be as cogent and may suffer under various idiosyncrasies. However, they will draw inspiration from other classic algebra texts such as Hungerford, Dummit and Foote, Jacobson's *Basic Algebra I* and M. Artin's *Algebra*.

Notable about B.H. is that it is unorthodox in pedagogy, viz. introducing rings before groups. These notes will follow this precedent, though perhaps of interest to note is that the classical treatment is groups → rings → fields (e.g. Hungerford's GTM *Algebra*).

[1] *Mathematics in the 20th Century. Bulletin of the London Mathematical Society. 2002.* Recommended reading (15 pages).

1. Chapter 1: Arithmetic in $\mathbb{Z}$

> *. . . for a student the content of a mathematical theory is never larger than the set of examples that are thoroughly understood.*
>
> – Vladimir Arnol'd[2]

Two good *models* for many ideas of abstract algebra are the integers

$$\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$$

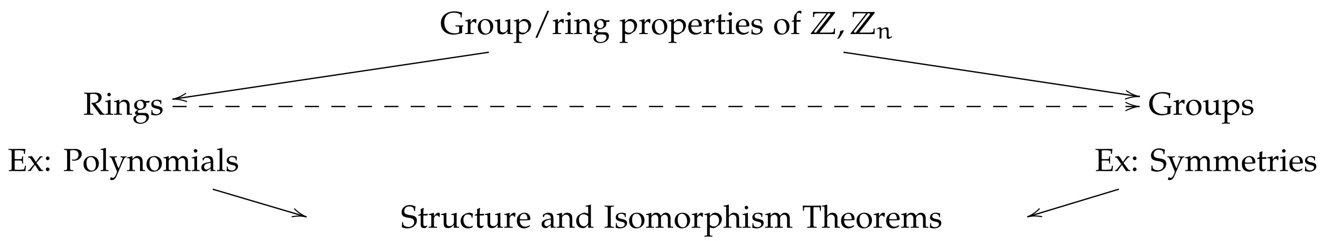
and 'the integers modulo n' denoted $\mathbb{Z}_n$

Remark 1.1. *What is a 'model'? We can see this word all over algebra textbooks. This is a loaded term. What is meant by 'model' is a specific instantiation of an idea - one that you keep sitting around in your head. When you learn new/abstract concepts about groups, rings, etc, you apply them to the model. That is a good way to get an intuition for how the new/abstract concept works.*

Page 'xvi' of B.H. has a thematic table of contents. The topics of this course are primes and factorization/decomposition, congruence, quotients and structure. We will study these first in $\mathbb{Z}$ and $\mathbb{Z}_n$ and then abstract to rings and groups. Chapters 1 and 2 of our book are mostly a review of these concepts. It is important to work with the concrete examples of these sections in order to develop a good intuition for the upcoming abstraction.

Remark 1.2. *What does it mean 'to abstract'? It is often the case that instances of ideas are first found in applications or are used implicitly. The idea is then recognized as important or interesting in its own right. Then an 'abstract definition' is made for the idea. This 'abstract definition' should capture and isolate what is unique/characteristic of the idea.*

$\mathbb{Z}, \mathbb{Z}_n$ work as concrete models for rings and groups. The archetypical example of a ring are the polynomials. For groups: symmetries of a shape[3]. A conceptual outline to keep in mind for the course may be as follows:


Group/ring properties of $\mathbb{Z}, \mathbb{Z}_n$
Rings
Groups
Ex: Polynomials
Ex: Symmetries
Structure and Isomorphism Theorems

1.1. Division Algorithm.

Proposition 1.3 (Well-Ordering Principle)**.** *Every nonempty subset of the set of nonnegative integers contains a smallest element.*

This is a set theoretic axiom.[4] This will be the workhorse of *many* of the proofs to come in this chapter.

[2] *Arnol'd, V. I. (2013). Lectures on partial differential equations. Springer Science & Business Media.*

[3] See S. Strogatz's article *Group Think*: https://opinionator.blogs.nytimes.com/2010/05/02/group-think/. Strogatz is a math professor at Cornell, and one of the best mathematical expositors.

[4] See https://en.wikipedia.org/wiki/Well-ordering_principle

Theorem 1.4 (Division Algorithm). *Let $a, b \in \mathbb{Z}$ be integers with $b > 0$. Then there exists unique integers q and r such that*

$$a = bq + r \qquad \textit{and} \qquad 0 \leqslant r < b$$

Remark 1.5. *We will apply the well-ordering principle. To set this up, we need a set of nonnegative integers. This proof is an archetypical application of the well-ordering principle!! This proof technique will be used over and over again!*

Proof. Let $a, b \in \mathbb{Z}$ be integers with $b > 0$. Let

$$S = \{a - bx : x \in \mathbb{Z} \text{ and } a - bx \geqslant 0\}$$

Step 1) Show that S is nonempty.

We must find a value for x such that $a - bx \geqslant 0$. What x can we choose? How about $x = -|a|$? Then check $a + b|a| \geqslant 0$

Step 2) Find q, r such that $a = bq + r$ and $r \geqslant 0$.

By the well-ordering principle S must contain a smallest element, call it r. Then $r = a - bq$ for some $q \in \mathbb{Z}$.

Step 3) Show that $0 \leqslant r < b$.

We have that $0 \leqslant r$ by defn of S. We will use a proof by contradiction to show that $r < b$. Suppose that $r \geqslant b$. Then

$$0 \leqslant r - b = (a - bq) - b = a - b(q + 1)$$

The right hand side is an element of S. Since $b > 0$ we have $a - b(q + 1) < a - bq = r$. This contradicts our choice of r.

Step 4) Show that r, q are unique. Suppose

$$a = bq_1 + r_1 \qquad a = bq_2 + r_2$$

with

$$0 \leqslant r_1 < b \qquad 0 \leqslant r_2 < b$$

We can multiply $0 \leqslant r_1 < b$ by -1 to get $-b < -r_1 \leqslant 0$. Adding these two equations we have $-b < r_2 - r_1 < b$.

We have $bq_1 + r_1 = a = bq_2 + r_2$. Rearranging, we have $r_2 - r_1 = b(q_1 - q_2)$

Therefore $-b < b(q_1 - q_2) < b$. Implying $-1 < q_1 - q_2 < 1$. Since $q_1 - q_2$ is an integer, it is forced to be zero. □

Remark 1.6. *Is it important that $b > 0$ as a hypothesis of the theorem? See Exercise 11 in 1.1.*

Remark 1.7. *Why is $0 \leqslant r < b$ important? If we do not specify this, then q, r need not be unique!*

Remark 1.8. *The division algorithm says something about the structure of $\mathbb{Z}$. This says something about how a breaks apart in terms of b: a is a multiple of b plus some remainder*

1.2. **Divisibility.** An important case of division occurs when for

$$a = bq + r$$

we have remainder $r = 0$.

Let $a, b \in \mathbb{Z}$ with $b \neq 0$. If $a = bc$ for some $c \in \mathbb{Z}$ then b is said to *divide* a, written as $b|a$. If b does not divide a, then we write $b \nmid a$.

An important concept concerning divisors is the following:

Definition 1.9 (Greatest Common Divisor). *Let* a *and* b *be integers, not both* 0*.* d *is* greatest common divisor *(gcd) of* a *and* b *if*

(1) $d|a$ *and* $d|b$
(2) *if* $c|a$ *and* $c|b$ *then* $c \leqslant d$

Denoted (a, b)

Theorem 1.10 (Bézout's identity). *Let* a *and* b *be integers, not both 0. Let* $d = (a, b)$*. There exists integers* $u, v \in \mathbb{Z}$ *such that* $d = au + bv$*.*

Remark 1.11. *This is another application of well ordering principle!*

Proof. Let

$$S = \{am + bn : m, n \in \mathbb{Z}, am + bn > 0\}$$

1) Show that S is nonempty.

Take $m = a, n = b$, then $a^2 + b^2 = aa + bb \geqslant 0$. Since a, b are not both zero, $a^2 + b^2 > 0$. Now apply the Well-Ordering Principle to get $t = au + bv$

2) We now wish to show $t = (a, b)$.

We must first show that $t|a$ and $t|b$. By the Division Algorithm,

$$a = tq + r, \qquad 0 \leqslant r < t$$

Now we can rewrite this as

$$r = a - tq = a - (au + bv)q = a - auq - bvq = a(1 - qu) + b(-vq)$$

Thus $r \in S$ and $r < t$. By our choice of t this forces $r = 0$. A similar argument shows that $t|b$.

3) We must show that t is greatest divisor. Let $c|a$ and $c|b$. Thus $a = kc$ and $b = k'c$. Then $t = au + bv = kcu + k'cv = c(ku + k'v)$. Then

$$|t| = |ku + k'v||c|$$

Thus $c \leqslant |c| \leqslant |t| = t$ (where the last follows as $t > 0$)

□

1.3. Primes and Factorization.

> *...[prime numbers] are fundamental. They're the atoms of arithmetic. Just as the Greek origin of the word 'atom' suggests the primes are 'a-tomic', meaning 'uncuttable, indivisible.' And just as everything is composed of atoms, every number is composed of primes.*
>
> – Steven Strogatz[5]

Primes are the building blocks of $\mathbb{Z}$. Primes determine structure (e.g. Fundamental Theorem of Arithmetic). Primes are 'atomic' or 'irreducible'.

Definition 1.12. *An integer* p *is* prime *if* $p \neq 0, \pm 1$ *and the only divisors of* P *are* ± 1 *and* $\pm p$*.*

Lemma 1.13. *Every integer* n *except* $0, \pm 1$ *is a product of primes.*

Remark 1.14. *Here product may mean a trivial product! Another application of the well-ordering principle! The idea of the proof is to let* S *be the set of all integers greater than* 1 *that are* **not** *a product of primes. Then show that* S *is the empty set. If* S *is empty every number is a product of primes!*

[5] The Loneliest Numbers. *The Joy of x*.

Proof. Note that if n is a product of primes, then so is $-n$. Thus we only need to demonstrate this for $n > 1$. Let

$$S = \{m \in \mathbb{Z} : m \geqslant 1 \text{ and } m \text{ is not a product of primes}\}$$

Suppose that S is not empty. Then by the Well Ordering Principle there is a smallest integer $m \in S$. Since m is not prime, there must be positive divisors $1 < a, b < m$. By our choice of m, we know $a, b \notin S$. Thus $a = p_1 p_2 \cdots p_r$ and $b = q_1 q_2 \cdots q_s$ for primes p_i, q_j. Thus

$$m = ab = p_1 p_2 \cdots p_r q_1 q_2 \cdots q_r$$

is a product of primes. So $m \notin S$. This is a contradiction of S being nonempty. Thus S is empty. □

Lemma 1.15 (Theorem 1.5 in B.H.). *Let p be an integer with $p \neq 0, \pm 1$. Then p is prime if and only if p has the following property:*

$$\textit{whenever } p|bc \textit{ then } p|b \textit{ or } p|c \tag{1}$$

Proof. (1) Let p be prime and assume $p|bc$. Consider $d = (b, p)$. Since $d|p$ either $d = 1$ or $d = p$.
Case 1) if $d = p$ then $p|b$.
Case 2) $d = 1$. By Bézout's Lemma there exist $u, v \in \mathbb{Z}$ with

$$1 = ub + vp$$

Multiplying by c on both sides:

$$c = (cb)u + cvp$$

Since $p|bc$ we have $bc = kp$ for some $k \in \mathbb{Z}$ thus

$$c = (kp)u + cvp = p(ku + cv)$$

(2) Suppose that p has property (1). We wish to show p is prime. Let $q|p$. Then $p = kq$. By hypothesis this implies $p|q$ or $p|k$.
Case 1) if $p|q$ then we have $p|q$ and $q|p$ thus $q = \pm p$.
Case 2) $p|k$ and $k|p$ thus $p = \pm k$ and $q = \pm 1$.

□

Corollary 1.16. *If p is prime and $p|a_1 a_2 \cdots a_n$ then $p|a_i$ for some i.*

Theorem 1.17 (Fundamental Theorem of Arithmetic). *Let $n \in \mathbb{Z}$ with $n \neq 0, \pm 1$. Then n is a product of primes. The prime factorization is unique, i.e. If $n = p_1 p_2 \cdots p_r$ and $n = q_1 q_2 \cdots q_s$ with each p_i, q_i prime, then $r = s$ and after relabeling the $q's$ $p_1 = \pm q_1, p_2 = \pm q_2, \ldots p_r = \pm q_r$*

Remark 1.18. *This is a decomposition theorem and a very important structure theorem. Informally, we can think of the primes forming a 'basis' for $\mathbb{Z}$!*

Proof. Suppose n has two prime factorizations, then

$$p_1(p_2 \cdot p_r) = q_1 q_2 \cdots q_s$$

Then $p_1 | q_1 q_2 \cdots q_s$ implies then $p_1 | q_j$ for some j by the Corollary. Without loss of generality (by relabeling or reordering), we may assume that $p_1 | q_1$. As p_1, q_1 are prime, we have $p_1 = \pm q_1$. Thus

$$\pm q_1 p_2 p_3 \cdots p_r = q_1 q_2 \cdots q_s$$

Divide both sides by q_1 to get that

$$p_2(\pm p_3 p_4 \cdots p_r) = q_2 q_3 \cdots q_s$$

We can then repeat the application of the Corollary.

Case 1) If $r = s$, then we are done.

Case 2) If $r \neq s$, then $r > s$ or $r < s$. If $r > s$ then we have $\pm p_{s+1} p_{s+2} \cdots p_r = 1$. This implies that $p_r | 1$. But as p_r is prime, this is a contradiction as the only divisors of 1 are ± 1. A similar contradction arises for $r < s$ (as this implies that $\pm 1 = q_r q_{r+1} \cdots q_s$ with the q_j's prime). Thus $r = s$.

□

.

Theorem 1.19. *Let $n > 1$. If n has no positive prime factor less than or equal to $\sqrt{n}$ then n is prime.*

Proof. Let $n > 1$. Suppose n is not prime. Then by the prime decomposition, we may write $n = p_1 p_2 k$ for primes p_1, p_2 and some integer $k \in \mathbb{Z}$. Without loss of generality we may assume that p_1, p_2, k are all positive integers. By hypothesis $p_1, p_2 > \sqrt{n}$. However this implies that

$$n = p_1 p_2 k \geqslant p_1 p_2 > \sqrt{n}\sqrt{n} = n$$

This is a contradiction. □

2. Chapter 2: Congruence

> *The invention of the symbol ≡ by Gauss affords a striking example of the advantages which may be derived from an appropriate notation, and marks an epoch in the development of the science of arithmetic.*
>
> – G.B. Mathews[6]

In this section we we'll study congruence classes of $\mathbb{Z}$. This definition dates back to Gauss.

'Congruence' is a notion of equivalence.

Congruence will be our model for quotient objects (See BH 'Congruence' in the thematic table of contents pg xvi)

2.1. Congruence Classes.

Consider the following simple observation: two $a, b \in \mathbb{Z}$ are equal $a = b$ if their difference is zero, i.e. $a - b = 0$, or a 'multiple' of zero $a - b = k0$

We say that two integers are *congruent modulo n* if $a - b = nk$ for some $k \in \mathbb{Z}$, i.e. there difference is a multiple of n.

Definition 2.1. $a, b, n \in \mathbb{Z}$ *with* $n > 0$. a *is congruent to* b *modulo* n, *written* $a \equiv b \mod n$ *if* $n|(a - b)$.

Congruence is an *equivalence relation*[7], meaning that

Theorem 2.2. *Let* $n > 0$. *Then for all* $a, b, c \in \mathbb{Z}$

(1) $a \equiv a \mod n$ *(reflexive)*
(2) *if* $a \equiv b \mod n$ *then* $b \equiv a \mod n$ *(symmetric)*
(3) *if* $a \equiv b \mod n$ *and* $b \equiv c \mod n$ *then* $a \equiv c \mod n$ *(transitive)*

Proof. (1) We must show $a = a \mod n$. We have that $n|0$ thus $n|(a - a)$.

(2) We want to show that $b - a = k'n$. By hypothesis

$$a - b = kn$$

for some $k \in \mathbb{Z}$. Mulitplying this equation by -1 we have

$$(b - a) = (-k)n$$

Thus $b \equiv a \mod n$.

(3) Let $a = b \mod n$ and $b = c \mod n$. We want to show $a = c \mod n$. By hypothesis $(a - b) = nk$ and $(b - c) = nk'$. Now, using the trick of adding zero,

$$a - c = a - b + b - c = (a - b) + (b - c) = nk + nk' = n(k + k')$$

□

Remark 2.3. *The fact that '$\equiv \mod n$' is an equivalence relation is fundamental - for a fixed* $n > 0$ $a \in \mathbb{Z}$ *has an equivalence class in* $\mathbb{Z}_n$, *we we'll call a 'congruence class' We next show that we can define addition/multiplication (group/ring) operations on the set of equivalence classes. This is a concrete example of what will be done in BH Chapter 6.2 (Quotient Rings) and 8.3,8.4 (Quotient Groups).*

[6] *Mathews, G. B. (1892). Theory of Numbers. Deighton Bell.*

[7] See https://en.wikipedia.org/wiki/Equivalence_relation

We now show how addition and multiplication work modulo n.

Theorem 2.4. *If* $a \equiv b \mod n$ *and* $c \equiv d \mod n$ *then*

1. $a + c \equiv b + d \mod n$
2. $ac \equiv bd \mod n$

Proof. By hypothesis we have $(a - b) = kn$ and $(c - d) = k'n$ for $k, k' \in \mathbb{Z}$.

1. We must show that $n|(a + c) - (b + d)$. We have
$$(a + c) - (b + d) = (a - b) + (c - d) = kn + k'n = (k + k')n$$
Thus $n|(a + c) - (b + d)$.
2. We must show $n|ac - bd$. Then
$$ac - bd = ac - ad + ad - bd = a(c - d) + d(a - b) = a(k'n) + d(kn) = n(k'a + kd)$$

□

Remark 2.5. *Note in the above proof of (2) the trick of 'adding zero'. This allows us to rewrite the equation - a very helpful proof technique.*

Definition 2.6. *Let* $a, n \in \mathbb{Z}$ *with* $n > 0$*. The congruence class of* a *modulo* n*, denoted* $[a]$ *is defined as follows:*
$$[a]_n = [a] = \{b \in \mathbb{Z} : b \equiv a \mod n\}$$

Sometimes the dependence on n is explicit in the notation, i.e. $[a]_n$. When n is clear from context, then it is written $[a]$.

We quick computation shows that
$$[a] = \{b : b \equiv a \mod n\} = \{b : b = a + kn \text{ for some } k \in \mathbb{Z}\} = \{a + kn : k \in \mathbb{Z}\}$$

Example (even/odd numbers): Let $n = 2$. Let's look at congruence classes modulo 2:
$$\begin{aligned}[2]_2 &= \{\ldots, 2 + (-2)2, 2 + (-1)2, 2 + 0, 2 + (1)2, 2 + (2)2, \ldots\}\\ &= \{\ldots, -2, \quad 0, \quad 2, \quad 4, \quad 6, \ldots\}\end{aligned}$$

and
$$\begin{aligned}[1]_2 &= \{\ldots, 1 + (-2)2, 1 + (-1)2, 1 + 0, 1 + (1)2, 1 + (2)2, \ldots\}\\ &= \{\ldots, -3, \quad -1, \quad 1, \quad 3, \quad 5, \ldots\}\end{aligned}$$

Note that $[0]_2 = [2]_2$ is the set of even integers (Why?)

Theorem 2.7 (Theorem 2.3 in B.H.)**.** $a \equiv c \mod n$ *if and only if* $[a] = [c]$

Proof. First, assume $a \equiv c \mod n$. We must show that $[a] = [c]$. To exhibit an equality we show both containments ($[a] \subset [c]$ and $[c] \subset [a]$).

We begin by showing that $[a] \subset [c]$. Let $b \in [a]$. Then $b = a + kn$ for $k \in \mathbb{Z}$. By hypothesis $a - c = k'n$, thus $a = c + k'n$. Thus
$$b = a + kn = (c + k'n) + kn = c + (k' + k)n$$

Therefore $b \in [c]$.

Since $a \equiv c \mod n$ symmetry implies $c \equiv a \mod n$, therefore the above argument may be applied to show that $[c] \subset [a]$.

Now assume that $[a] = [c]$. We wish to show that $a \equiv c \mod n$. Since $a \in [a] = [c]$, we have $a = c + kn$ for some $k \in \mathbb{Z}$. Thus $n|(a - c)$. □

Remark 2.8. *Let's reinterpret Theorem 2.4. It says that if* $[a] = [b]$ *and* $[c] = [d]$ *then* $[a + c] = [b + d]$ *and* $[ac] = [bd]$*. Thus addition and multiplication of integers preserve congruence classes.*

One of the powerful aspects of congruence classes (or equivalence classes) is that they partition the set.

Corollary 2.9. *Two congruence classes are either disjoint or identical.*

Proof. Let $[a]$ and $[c]$ be congruence classes. If $[a] \cap [c] = \emptyset$ then they are disjoint. If not, then there exists $b \in [a] \cap [c]$. Thus $b \equiv a \mod n$ and $b \equiv c \mod n$. By Theorem 2.3 this implies $[a] = [b] = [c]$.

□

Corollary 2.10. *Let* $n > 1$*.*

(1) *If* $a \in \mathbb{Z}$ *and* $a = nq + r$ *with* $0 \leqslant r < n$ *then* $[a] = [r]$
(2) *There are exactly* n *distinct congruence classes, namely* $[0], [1], \dots, [n - 1]$

Proof. (1) Assume $a = nq + r$ with $0 \leqslant r < n$. Thus $a - r = qn$. So $a = r \mod n$. Thus $[a] = [r]$.

(2) Let $a \in \mathbb{Z}$. Since

$$a = nq + r, \qquad 0 \leqslant r$$

by the Division, Algorithm, Part 1 implies that $[a] = [r]$ for some $0 \leqslant r < n$.

Thus $[a]$ must be one of $[0], [1], \dots [n - 1]$. Therefore we've shown there are at most n equivalence classes.

We must now show that these n classes are distinct. We do this by showing no two of $0, 1, 2, \dots, n - 1$ are congruent modulo n. Without loss of generality let $0 \leqslant s < t \leqslant n - 1$. Thus $0 \leqslant t - s \leqslant n - 1$ Hence $n \nmid t - s$, thus $t \not\equiv s \mod n$. Therefore by Corollary 2.9 $[t] \neq [s]$. Thus these are all distinct.

□

Definition 2.11. *The set of congruence classes modulo* n *is denoted* $\mathbb{Z}_n$

Remark 2.12. *The elements of* $\mathbb{Z}_n$ *are equivalence classes, not integers. Thus the 'modular arithmetic' done on* $\mathbb{Z}_n$ *will be defined on these equivalence classes.*

2.2. **Modular Arithmetic.** The set $\mathbb{Z}_n$ is the set of congruence classes. Since it is so closely related to $\mathbb{Z}$ - a natural question is ask is: are there operations on $\mathbb{Z}_n$?

To define addition, we must define what it means to 'add' two congruence classes. Since we have notions of addition and multiplication in $\mathbb{Z}$, we can 'define'

$$[a] + [c] := [a + c]$$

Is this well defined?

Similarly, we can 'define'

$$[a][c] := [ac]$$

Is this well-defined?

Remark 2.13. *What does it mean to be 'well-defined'? It implies that the result is independent of the representative, i.e. if* $[a] = [b]$ *and* $[d] = [c]$ *is* $[a] + [c] = [b] + [d]$*?*

Theorem 2.14. *If* $[a] = [b]$ *and* $[c] = [d]$ *in* $\mathbb{Z}_n$ *then*

$$[a + c] = [b + d] \quad \textit{and} \quad [ac] = [bd]$$

Proof. Since $[a] = [b]$ and $[c] = [d]$ we have that $a \equiv b \mod n$ and $c \equiv d \mod n$ by Theorem 2.7. We can now invoke Theorem 2.4 to get

$$a + c \equiv b + d \mod n$$

and

$$ac = bd \mod n$$

Invoking Theorem 2.7 again gives us $[a + c] = [b + d]$ and $[ac] = [bd]$.

□

Therefore we have well defined arithmetic in $\mathbb{Z}_n$ as

$$[a] + [c] = [a + c] \qquad [a][c] = [ac]$$

Remark 2.15. *Be careful here. The addition in the expression* $[a] + [c]$ *is addition on congruence classes. The addition in the expression* $[a + c]$ *is addition in* $\mathbb{Z}$*. The same with multiplication.*

Here are some properties:

Proposition 2.16. *For* $[a], [b], [c] \in \mathbb{Z}_n$ *we have*

(1) *if* $[a], [b] \in \mathbb{Z}_n$ *then* $[a] + [b] \in \mathbb{Z}_n$ *(closure of addition)*
(2) $[a] + ([b] + [c]) = ([a] + [b]) + [c]$ *(Associativity of addition)*
(3) $[a] + [b] = [b] + [a]$ *(Commutativity of addition)*
(4) $[a] + [0] = [a] = [0] + [a]$ *(Identity element for add.)*
(5) *For* $[a] \in \mathbb{Z}_n$ *there exists* $[b] \in \mathbb{Z}_n$ *such that* $[a] + [b] = [0]$ *(additive inverse)*

(6) *If* $[a], [b] \in \mathbb{Z}_n$ *then* $[a][b] \in \mathbb{Z}_n$ *(closure of multiplication)*
(7) $[a]([b][c]) = ([a][b])[c]$ *(assoc. of mult.)*
(8) $[a][b] = [b][a]$ *(commutativity of mult.)*
(9) $[a][1] = [a] = [1][a]$ *(identity element for mult.)*

(10) $[a]([b] + [c]) = [a][b] + [a][c]$ *and* $([a] + [b])[c] = [a][c] + [b][c]$ *(distributivity)*

Remark 2.17. $(1) - (5)$ *say that* $+$ *is an abelian group.* $(6) - (9)$ *say that mult. is a commutative monoid. Distributivity is a ring axiom that links them together.*

Proof. $(1), (6)$ follow straight from definitions. For (2)

$$[a] + ([b] + [c]) = [a] + [b + c] = [a + (b + c)] = [(a + b) + c] = [a + b] + [c] = ([a] + [b]) + [c]$$

The proofs of the others follow from using the definition and similar operations in $\mathbb{Z}$.

□

Remark 2.18. *The above section says that since addition and multiplication preserve congruence, they can be used to define 'addition' and 'multiplication' of congruence classes themselves. This means that the set of congruence classes (modulo* n*) form a ring (the subject of chapter 3). The map*

$$\mathbb{Z} \to \mathbb{Z}_n$$

sending an integer $a \to [a]$ *is compatible with addition and multiplication (it is a ring homomorphism).*

Example: Let $n = 13$. Let's compute $([7]+[9])([11]+[16])$. Notice we could do lift this to the integers and compute $(7+9)(11+16) \mod 3$. Instead, we can use our rules, i.e.

$$([7]+[9])([11]+[16]) = ([7+9])([11+16]) = [3][4] = [12]$$

2.3. **Structure of $\mathbb{Z}_n$.** The structure of $\mathbb{Z}_n$ - particularly when n is prime - is of great importance.[8]

Theorem 2.19. *Let $p > 1$. Then the following are equivalent (TFAE):*

(1) p *is prime*
(2) *For any $[a] \neq [0]$ (in $\mathbb{Z}_p$) the equation $[a][x] = [1]$ has a solution in $\mathbb{Z}_p$*
(3) *Whenever $[b][c] = 0$ (in $\mathbb{Z}_p$) then $[b] = [0]$ or $[c] = [0]$*

Proof. We show $(1) \implies (2)$. Assume $p \in \mathbb{Z}$ is prime and $[a] \neq [0] \in \mathbb{Z}_p$. Thus $a \neq 0 \mod p$, and $p \not| a$. Let $d = (a,p)$. Then $d = 1$ or $d = p$. But $d \neq p$ since $d|a$ and $p \not| a$. Thus we have $d = 1$.

By Bezout's Lemma we may write $1 = ua + vp$ for $u, v \in \mathbb{Z}$. Thus $au - 1 = (-v)p$. Thus $p|au-1$, implying $au = 1 \mod p$, i.e. $[au] = [1]$. Therefore $[a][u] = [1]$.

$(2) \implies (3)$. Suppose $[b][c] = 0$ in $\mathbb{Z}_p$. We wish to show that $[b] = [0]$ or $[c] = [0]$ (in $\mathbb{Z}_p$). If $[b] = [0]$ then we're done. If $[b] \neq 0$, then by hypothesis there exists $[u]$ such that

$$[b][u] = [1]$$

Multiplying this equation by $[c]$ we have

$$0 = [cb][u] = [c][b][u] = [c][1] = [c]$$

$(3) \implies (1)$. Assume that b, c are any integers such that $p|bc$. Thus $bc = 0 \mod p$, implying $[b][c] = [bc] = [0]$ in $\mathbb{Z}_p$. Therefore by hypothesis either $[b] = 0$ or $[c] = 0$. This implies $b = 0 \mod p$ or $c = 0 \mod p$, implying $p|b$ or $p|c$. Thus by a previous theorem (Theorem 1.5 in BH) p is prime. □

Remark 2.20. *Recall the structure of 'TFAE' proofs - see page 508 in BH.*

Theorem 2.21. *Let $a, n \in \mathbb{Z}$ with $n > 1$. Then $[a]x = [1]$ has a solution in $\mathbb{Z}_n$ if and only if $(a,n) = 1$ in $\mathbb{Z}$.*

Proof. This proof is an if and only if. Thus there are two parts.

Assume that the equation has a solution. We will show that $(a,n) = 1$. If $[a][w] = [1]$, then

$$\begin{aligned} [a][w] &= [1] && (2)\\ [aw] &= [1] && (3)\\ aw &\equiv 1 \mod n && (4)\\ aw - 1 &= kn \text{ for some } k \in \mathbb{Z} && (5)\\ aw + n(-k) &= 1 && (6) \end{aligned}$$

Let $d = (a,n)$. Then $dr = a$ and $ds = n$ for some $r, s \in \mathbb{Z}$. So

$$1 = aw + n(-k) = (dr)w + ds(-k) = d(rw - sk)$$

[8] See Theorems 9.7-9.12 in BH to see how these $\mathbb{Z}_p$ will be fundamental building blocks of any finite abelian group.

Thus $d|1$, however as $d > 0$ we have $d = 1$.

Now assume $(a, n) = 1$. We must show $[a]x = [1]$ has a solution in $\mathbb{Z}_n$. By Bezout's Lemma there exists integers $u, v \in \mathbb{Z}$ such that

$$1 = ua + vn$$

Rearranging, $ua - 1 = (-v)n$. Therefore $ua = 1 \mod n$ and

$$[u][a] = [ua] = [1]$$

□

An element $[a]$ in $\mathbb{Z}_n$ is called a **unit** if $[a][x] = [1]$ has a solution. In this case there there exists $[b]$ with $[a][b] = [1]$ and we say that $[b]$ is the *inverse* of $[a]$. Restating this last theorem in terms of units:

Theorem 2.22. *Let a and n be integers with $n > 1$. Then $[a]$ is a unit in $\mathbb{Z}_n$ if and only if $(a, n) = 1$ in $\mathbb{Z}$*

---

3. Chapter 3: Ring Theory

The axiomatic method has many advantages over honest work.

– Bertrand Russell[9]

A ring is a set on which you have two operations: 'addition' and 'multiplication'. We'll show that there is a notion of 'subtraction' in all rings (additive inverse). However, we will observe that in some rings there is not a notion of 'division' (multiplicative inverse).

3.1. **Definition and Examples.**

We'll start with familiar examples. Then we'll introduce the axioms which formalize their common properties. This allows us to prove theorems for 'arbitrary rings' (meaning any example that obeys the ring axioms). Therefore the results we prove will be valid for our specific examples. This is the power and process of *abstract* algebra.

3.1.1. *Number Systems.* $\mathbb{Z}, \mathbb{Z}_n$ are the rings we have just studied in Chapter 1,2. The familiar sets $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are examples of rings.

3.1.2. *Matrices.* Let $M_2(\mathbb{R})$ be the set of all 2×2 matrices over $\mathbb{R}$, i.e.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad a, b, c, d \in \mathbb{R}$$

Recall that addition is defined component-wise:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} a + a' & b + b' \\ c + c' & d + d' \end{pmatrix}$$

And multiplication is defined as follows:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} w & x \\ y & z \end{pmatrix} = \begin{pmatrix} aw + by & ax + bz \\ cw + dy & cx + dz \end{pmatrix}$$

3.1.3. *Functions.* Let $\mathcal{F} := \{f : \mathbb{R} \to \mathbb{R}\}$. Then we may define addition and multiplication component-wise as

$$(f + g)(x) = f(x) + g(x) \qquad (fg)(x) = f(x)g(x)$$

Are there zero divisors? i.e. $f, g \neq 0$ such that $fg = 0$?

3.1.4. *Axiomitize.* Let us now axiomitize the common properties of these sets.

Remark 3.1. *By 'axiomitize' we mean we will write down explicit axioms (1-8) that these sets obey which capture the common characteristics that we want to isolate. In particular,* $(R, +)$ *is an abelian group,* $(R, \cdot)$ *is a semigroup and there is a relationship between* $+, \cdot$ *given by distributivity.*

This should look similar to Theorem 2.7 in BH where we wrote down properties of multiplication and addition in $\mathbb{Z}_n$.

Definition 3.2 (Ring)**.** A *ring* is a nonempty set R equipped with two operations, $(+, \cdot)$ that satisfy the following:

(1) If $a \in R$ and $b \in R$ then $a + b \in R$ (Closure of addition)

[9] An oft-quoted paraphrase of *Russell, B. (1920). Introduction to Mathematical Philosophy.*

(2) $a + (b + c) = (a + b) + c$ (associative addition)
(3) $a + b = b + a$ (commutativity of addition)
(4) There exists 0_R in R such that $a + 0_R = a = 0_R + a$ for every $a \in R$ (Additive identity element)
(5) For each $a \in R$ there exists $x \in R$ such that $a + x = 0_R$ (additive inverse element)
(6) If $a \in R$ and $b \in R$ then $ab \in R$ (closure of multiplication)
(7) $a(bc) = (ab)c$ (Associative multiplication)
(8) $a(b + c) = ab + ac$ and $(a + b)c = ac + bc$ (Distributive laws)

Remark 3.3. *Conditions* $(1) - (5)$ *stipulate constraints on the* $+$ *operation - in fact, they indicate that* $(R, +)$ *is an* **abelian group**. *Conditions* $6, 7$ *constrain the* $\cdot$ *operation, they show that* $(R, \cdot)$ *is a* **semigroup**.

Rings do not necessarily have a multiplicative identity!

Definition 3.4. A *ring with identity* is a ring R that contains an element 1_R such that $a1_R = a = 1_Ra$ for all $a \in R$

Remark 3.5. *Such rings are also called* **unital rings**. *Be very careful if you are looking at the statements about rings in different texts (e.g. wikipedia). Some assume rings have a multiplicative identity and some do not. Our text does not assume it. This becomes most salient when you are trying to prove various statements about rings!*[10]

Remark 3.6. *The standard example of a ring without multiplicative identity is* $2\mathbb{Z} := \{2n : n \in \mathbb{Z}\}$*, i.e. the even integers. Other examples of rings without multiplicative identities come from analysis (for instance some particular ring of functions not containing* $f(x) \equiv 1$*). Can you find any? Here's one.* $f : \mathbb{R} \to \mathbb{R}$ *is said to have* compact support *if there exist are real numbers* a, b *(depending on fo such that* $f(x) = 0$ *for* $x \notin [a, b]$*, (i.e.* f *is zero outside a bounded interval). The set of functions* $f : \mathbb{R} \to \mathbb{R}$ *with compact support is a commutative ring without identity - Why?). However,as most of our concrete examples in this class will come from algebra, they will typically have multiplicative identities.*

Moreover, the multiplication is not necessarily commutative.

Definition 3.7. A *commutative ring* is a ring R that such that $ab = ba$ for all $a, b \in R$.

Which of the examples above is not commutative? Matrices are the one of the best models for noncommutative rings.

3.1.5. *More Concepts.* We've seen that nonzero elements can multiply to give zero. In many cases this is undesirable and we can strengthen the definition by excluding this:

Definition 3.8 (Integral Domain)**.** An *integral domain* is a commutative ring R with identity $1_R \neq 0_R$ such that if $a, b \in R$ and $ab = 0_R$ then $a = 0_R$ or $b = 0_R$

Remark 3.9. *Another way to say this is that an integral domain is a commutative ring that contains no zero divisors!*

Remark 3.10. *The condition* $1_R \neq 0_R$ *excludes the* 0 *ring from being an integral domain.*

Is $\mathbb{Z}$ an integral domain? $\mathbb{Z}_n$? For what types of n?

[10] See https://en.wikipedia.org/wiki/Rng_(algebra) for some discussion.

Definition 3.11 (Field)**.** A *field* is a commutative ring R with identity $1_R \neq 0_R$ such that for each $a \in R$ with $a \neq 0_R$ the equation $ax = 1_R$ has a solution in R.

What are some examples? How about $\mathbb{Q}$, $\mathbb{R}$? $\mathbb{Z}_p$?

3.1.6. *Subrings.*

Definition 3.12. Let $S \subset R$. If S is a ring then S is called a *subring* of R.

Examples of subrings:
Let $C(\mathbb{R}) := \{f \in \mathcal{F} : f \text{ continuous}\} \subset \mathcal{F}$. Is this a subring? Why? From our knowledge of continuous functions this is a subring.
Conditions of a subring: checking some subset $S \subset R$ is a subring is easier than checking that S is a ring.

Theorem 3.13. *Let R be a ring. Let $S \subset R$. Suppose that*

(1) *S is closed under addition*
(2) *S is closed under multiplication*
(3) $0_R \in S$
(4) *if $a \in S$ then $a + x = 0$ has solution in S (closed under inverses)*

then S is a subring of R.

Why? If R is a ring, then elements of S already obey the other properties.
Example Applications:

$$GL_2(\mathbb{R}) = \{A \in M_2(\mathbb{R}) : \text{ there exists } B \in M_2(R) \text{ such that } AB = 1_{M_2} = BA\}$$

Is $GL_2(\mathbb{R})$ a subring of $M_2(\mathbb{R})$? (Why not? How about the (3) in Theorem 3.13?)

Proposition 3.14. *Let $\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Z}\}$. Show this a subring of $\mathbb{R}$.*

Proof. We'll apply the previous theorem.

(1) $a + b\sqrt{2} + c + d\sqrt{2} = (a + c) + (b + d)\sqrt{2} \in \mathbb{Z}[\sqrt{2}]$.
(2) $(a + b\sqrt{2})(c + d\sqrt{2}) = ac + ad\sqrt{2} + bc\sqrt{2} + 2bd = (ac + 2bd) + (ad + bc)\sqrt{2} \in \mathbb{Z}[\sqrt{2}]$
(3) $0 = 0 + 0\sqrt{2} \in \mathbb{Z}[\sqrt{2}]$
(4) For $a + b\sqrt{2}$ we have $(a + b\sqrt{2}) + (-a + -b\sqrt{2}) = 0$.

□

3.2. Properties of Rings.

3.2.1. *Arithmetic in Rings.* Everyone is familiar with the addition and 'subtraction' in the ring $\mathbb{Z}$. However subtraction is not explicit in the definition of a ring. By

$$a - b$$

we mean $a + (-b)$, where $-b$ is the solution to $b + x = 0$. We now show this solution is unique.

Theorem 3.15. *Let R be a ring. For $a \in R$ the equation $a + x = 0$ has a unique solution.*

Proof. By definition of ring, $a + x = 0$ has some solution u. Suppose u, v are both solutions. Then

$$v = 0 + v = (a + u) + v = (u + a) + v) = u + (a + v) = u + 0 = u$$

Thus u is unique. □

Remark 3.16. *It is important to notice that '$-$' is not an operation! It is notation for an inverse element, i.e. the notation $-a$ denotes the unique element of $-a \in R$ such that*

$$a + (-a) = 0 = (-a) + a$$

Consider the equation

$$a + b = a + c$$

with a, b, c elements of an arbitrary ring R. Does this imply $b = c$? In $\mathbb{Z}$ we know it does. For an arbitrary ring we need to know that we have a 'cancellation law' meaning that we can add the element '-a' to both sides to get $b = c$

Theorem 3.17. *If $a + b = a + c$ in R then $b = c$*

Proof.

$$\begin{aligned} a + b &= a + c \\ -a + (a + b) &= -a + (a + c) \qquad \text{add } -a \text{ to both sides} \\ (-a + a) + b &= (-a + a) + c \qquad \text{Use associativity of } + \\ 0_R + b &= 0_R + c \qquad \text{use additive identity, i.e. Axiom 4} \\ b &= c \end{aligned}$$

□

More properties than are familiar from $\mathbb{Z}$ and hold for arbitrary rings:

Theorem 3.18. *Let R be a ring. Let $a, b \in R$. Then*

(1) $a0_R = 0_R = 0_R a$

Proof. $a(0_R) = a(a - a) = aa - aa = 0$ and $0_R a = (a - a)a = aa - aa = 0_R$. □

(2) $a(-b) = -ab$ *and* $(-a)b = -ab$

Proof. We must show that $a(-b)$ solves the equation $ab + x = 0_R$. $ab + a(-b) = a(b - b) = 0_R$ □

(3) $-(-a) = a$

Proof. $-(-a)$ is the solution to $-a + x = 0_R$. Since a solves this equation, by uniqueness $a = -(-a)$. □

(4) $-(a + b) = (-a) + (-b)$

Proof. $(a + b) + (-a + -b) = a + (-a) + b + (-b) = 0_R + 0_R = 0_R$ □

(5) $-(a - b) = -a + b$

Proof. $-(a - b) = -(a + (-b)) = (-a) + -(-b) = -a + b$ □

(6) $(-a)(-b) = ab$

Proof. By using two equations of (2) we have $(-a)(-b) = -(a(-b)) = -(-ab) = ab$ □

and if R *has an identity, then*

(7) $(-1_R)a = -a$

Proof. $(-1_R)a = -(1_R a) = -(a) = -a$ □

Subtraction provides an easier theorem for showing that a subset is a subring.

Theorem 3.19. *Let* R *be a ring. Let* $S \subset R$ *such that* $S \neq \emptyset$ *and*

(1) *For* $a, b \in S$ *we have* $a - b \in S$ *(S closed under subtraction)*
(2) *For* $a, b \in S$ *we have* $ab \in S$ *(closed under multiplication)*

Proof. We show that S satisfies the conditions of our previous theorem.

(1) We must show S is closed under multiplication. This is Part 2 of our hypothesis.
(2) We must show that $0_R \in S$. Since S is nonempty, there is some $c \in S$. By hypothesis $0 = c - c \in S$. (Closure of subtraction)
(3) We must show that if $a \in S$ then $-a \in S$. Since we have $0 \in S$ we have $-a = 0_R - a \in S$.
(4) We must show closure of addition. Let $a, b \in S$. Since we have $-b \in S$ we can write $a + b = a - (-b) \in S$.

□

3.3. **Units, Zero Divisors.** We introduced units and zero divisors in $\mathbb{Z}_n$. We now explore these in arbitrary rings.

Definition 3.20 (Field)**.** Let R be a ring with identity. $a \in R$ is a *unit* if there exists $u \in R$ such that

$$au = 1_R = ua$$

In this case u is called the *multiplicative inverse* of a and is denoted a^{-1}.

What are the units of $\mathbb{Z}$? (± 1)

Here's a relationship between fields and units. Let F be a field. Then F is commutative ring with identity, and by definition there exists some $u \in F$ such that $au = 1_R$. Thus every element of a field is a unit.

Definition 3.21 (Zero Divisor)**.** Let R be a ring. Then $a \in R$ is a *zero divisor* if

(1) $a \neq 0_R$
(2) There exists a nonzero element $c \in R$ such that $ac = 0_R$ or $ca = 0_R$

Note that c is not necessarily unique. Consider $\mathbb{Z}_6$. For $[3] \in \mathbb{Z}_6$ we have $[2][3] = 0$ and $[4][3] = 0$ with $[2] \neq [4]$.

And for a noncommutative ring we may have $ac = 0_R$ and $ca \neq 0_R$ Can you find an example? What's the archetypical noncommutative ring? Here's an example in $M_2(R)$:

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

What's the relationship between integral domains and zero divisors? An integral domain satisfies the property that if $ab = 0$ then $a = 0$ or $b = 0$. The contrapositive of this is $a \neq 0, b \neq 0$ implies $ab \neq 0$. Thus integral domains cannot contain zero divisors.

Remark 3.22. *Integral domains are important because they imply a cancellation property for the multiplication!*

Theorem 3.23. *Cancellation is valid in any integral domain R, i.e. if $a \neq 0_R$ and $ab = ac$ in R then $b = c$*

Proof. Let $ab = ac$. Then $ab - ac = 0$. Thus $a(b - c) = 0$. Since $a \neq 0_R$ then $b - c = 0_R$ (Why? if not a would be a zero divisor). Thus $b = c$. □

Theorem 3.24. *Let F be a field. Then F is an integral domain.*

Proof. We must show that if $ab = 0_F$ then $a = 0_F$ or $b = 0_F$. Let $a, b \in F$ such that

$$ab = 0$$

If $b = 0$ then we're done. If $b \neq 0$, then b is a unit is F is a field. Thus applying b^{-1} to the equation we have

$$a = abb^{-1} = 0b^{-1} = 0$$

Thus $a = 0$. Therefore F is an integral domain. □

Remark 3.25. *The converse of this theorem (i.e. an integral domain is a field) is not true in general! For instance, consider $\mathbb{Z}$. However, it is true in the finite case.*

Remark 3.26. *The next theorem is very relevant to the workshop problems. In fact, is an abstract version of one of the more concrete workshop problems. Can you see the resemblance?*

Theorem 3.27. *Let R be a finite integral domain. Then R is a field.*

Proof. Let R be a finite integral domain. Then R is a finite commutative ring with identity. Since R is finite, $|R| = n$. Thus we only need to show that for $a \neq 0_R$ the equation $ax = 1_R$ has a solution. Let $a \in R$ with $a \neq 0_R$.

Consider the set $S = \{ax : x \in R\} \subseteq R$. What is its cardinality? Suppose $|S| < n$. Then we must have $ax = ay$ for $x, y \in R$ with $x \neq y$. Thus $a(x - y) = 0_R$. This implies $x = y$ since R is an integral domain. This is a contradiction.

Thus $|S| = n$. This implies $S = R$. Therefore $1_R \in S$ and there exists $y \in R$ such that $ay = 1_R$. Thus R is a field.

□

A relationship to keep in mind is the following:

$$\text{Fields} \subseteq \text{Integral Domains} \subset \text{Rings}$$

The first $\subseteq$ comes from Theorem 3.24. Moreover for finite rings, we have shown that this inclusion is equality: Fields = Integral Domains.

3.4. Morphisms.

> *. . . much of Mathematics is dynamic, in that it deals with morphisms of an object into another object of the same kind. Such morphisms (like functions) form categories, and so the approach via categories fits well with the objective of organizing and understanding Mathematics.*
>
> – Saunders MacLane[11]

Isomorphic rings are rings that have the same structure.

Here is the intuitive idea: rings R and S are isomorphic if one can relabel the elements of R to get S. Let's look at an example.

Let $S = \{0, 5\} \subset \mathbb{Z}_{10}$. Is S a subring? (Show that it is closed under subtraction and multiplication)

We claim that S has 'the same structure' as $\mathbb{Z}_2$. Speaking coarsely, what is meant by this is that up to a relabeling the multipication and addition tables are the same. Therefore the operations $(+, \cdot)$ in S work in the same way as those in $\mathbb{Z}_2$, i.e. the ring S 'is' $\mathbb{Z}_2$ with different labels.

In S we have the following addition/multiplication tables. Recall that $0 = [0] \in \mathbb{Z}_{10}$ and $5 = [5] \in \mathbb{Z}_{10}$.

$$\begin{array}{c|cc} + & [0]_{10} & [5]_{10} \\ \hline [0]_{10} & 0 & 5 \\ [5]_{10} & 5 & 0 \end{array} \qquad\qquad \begin{array}{c|cc} \cdot & [0]_{10} & [5]_{10} \\ \hline [0]_{10} & 0 & 0 \\ [5]_{10} & 0 & 5 \end{array}$$

In $\mathbb{Z}_2$ we have the following tables. Recall that $0 = [0] \in \mathbb{Z}_2$ and $1 = [1] \in \mathbb{Z}_2$.

$$\begin{array}{c|cc} + & [0]_{2} & [1]_{2} \\ \hline [0]_{2} & 0 & 1 \\ [1]_{2} & 1 & 0 \end{array} \qquad\qquad \begin{array}{c|cc} \cdot & [0]_{2} & [1]_{2} \\ \hline [0]_{2} & 0 & 0 \\ [1]_{2} & 0 & 1 \end{array}$$

Consider the relabeling

$$[0]_{10} \mapsto [0]_2 \qquad [5]_{10} \mapsto [1]_2$$

This is a function from $S \to \mathbb{Z}_2$ that transforms the multiplication/addition tables in S to those of $\mathbb{Z}_2$.

Definition 3.28. A ring R is *isomorphic* to a ring S if there is a function $f : R \to S$ such that

(1) f is injective
(2) f is surjective
(3) $f(a + b) = f(a) + f(b)$ $\qquad f(ab) = f(a)f(b)$ for all $a, b \in R$

In this case the function f is called an *isomorphism*.

3.4.1. *Example: Complex Numbers.* Consider the set K of 2×2 matrices of the form

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix} \qquad a, b \in \mathbb{R}$$

[11] *MacLane, S. (2012). Mathematics form and function. Springer Science & Business Media.* Historical anecdote: Saunders MacLane was the PhD advisor of Thomas W. Hungerford, the author of BH https://en.wikipedia.org/wiki/Thomas_W._Hungerford

We claim that K is isomorphic to the field $\mathbb{C}$ of complex numbers. To show this, define the function $f : K \to \mathbb{C}$ by

$$f\begin{pmatrix} a & b \\ -b & a \end{pmatrix} = a + bi$$

We show that f is injective, assume that

$$a + bi = f\begin{pmatrix} a & b \\ -b & a \end{pmatrix} = f\begin{pmatrix} r & s \\ -s & r \end{pmatrix} = r + si$$

Then $a = r$ and $s = b$. Therefore

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix} = \begin{pmatrix} r & s \\ -s & r \end{pmatrix}$$

f is surjective, as for any $a + bi \in \mathbb{C}$ we have

$$f\begin{pmatrix} a & b \\ -b & a \end{pmatrix} = a + bi$$

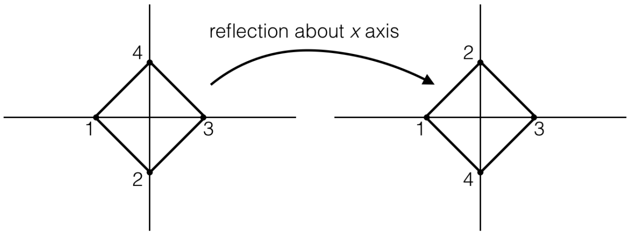
Finally, we have

$$f\left[\begin{pmatrix} a & b \\ -b & a \end{pmatrix} + \begin{pmatrix} c & d \\ -d & c \end{pmatrix}\right] = f\begin{pmatrix} a+c & b+d \\ -b-d & a+c \end{pmatrix} = (a+c) + (b+d)i$$
$$= (a + bi) + (c + di) = f\begin{pmatrix} a & b \\ -b & a \end{pmatrix} + f\begin{pmatrix} c & d \\ -d & c \end{pmatrix}$$

and

$$f\left[\begin{pmatrix} a & b \\ -b & a \end{pmatrix}\begin{pmatrix} c & d \\ -d & c \end{pmatrix}\right] = \begin{pmatrix} ac-bd & ad+bc \\ -ad-bc & ac-bd \end{pmatrix} = (ac - bd) + (ad + bc)i$$
$$= (a + bi)(c + di) = f\begin{pmatrix} a & b \\ -b & a \end{pmatrix} f\begin{pmatrix} c & d \\ -d & c \end{pmatrix}$$

3.4.2. *Example: Reflection.* Let $f : \mathbb{C} \to \mathbb{C}$ be the complex conjugation map given by $f(a + bi) = a - bi$. This function has a geometric interpretation in the complex plane, where $a + bi$ is identified with the point (a, b).



The function f reflects the plane over the real line $\mathbb{R}$, with $(a, b) \mapsto (a, -b)$.[12]

The function f satisfies:

$$f\big[(a + bi) + (c + di)\big] = f\big[(a + c) + (b + d)i\big] = (a + c) - (b + d)i$$
$$= (a - bi) + (c - di) = f(a + bi) + f(c + di)$$

[12] This reflection is an example of a *symmetry*. Symmetries are often formalized as distance/structure-preserving bijections from an object to itself. We'll study symmetries more when we get to group theory.

and

$$\begin{aligned} f\big[(a+bi)(c+di)\big] = f\big[(ac-bd)+(ad+bc)i\big] &= (ac-bd)-(ad+bc)i \\ &= (a-bi)(c-di) = f(a+bi)f(c+di) \end{aligned}$$

Isomorphic rings are an important example. But more important are morphisms that preserve 'structure' meaning they respect the multiplication/addition of both rings.

Definition 3.29. Let R, S be rings. A function $f : R \to S$ is said to be a *ring homomorphism* if

(1) $f(a+b) = f(a)+f(b)$ (f is additive)
(2) $f(ab) = f(a)f(b)$ (f is multiplicative)

for all $a, b \in R$

Remark 3.30. *f is called 'additive' if f obeys condition (1). f is called 'multiplicative' if f obeys condition (2).*

A homomorphism which is injective is called a *monomorphism*. A surjective homomorphism is called an *epimorphism*. As we have seen a homomorphism which is injective and surjective is an *isomorphism*. For a fixed $b \in S$ the set $f^{-1}(b) = \{a \in R : f(a) = b\}$ is called the *fiber over* b.

3.4.3. *Examples.*

(1) For rings R and S, the *zero map* $f : R \to S$ is given by $f(r) = 0_S$ for all $r \in R$. We write $f \equiv 0$. Why is this a homomorphism?

$$f(a+b) = 0_S = 0_S + 0_S = f(a)+f(b) \qquad f(ab) = 0_S = 0_S 0_S = f(a)f(b)$$

(2) Consider the map $f : \mathbb{Z} \to \mathbb{Z}_2$, defined by

$$\mathbb{Z} \ni n \mapsto [n] \in \mathbb{Z}_2$$

The map is additive and multiplicative as

$$f(n+m) = [n+m] = [n]+[m] = f(n)+f(m) \qquad f(nm) = [nm] = [n][m] = f(n)f(m)$$

Notice that if n is even that $[n] = [0]$. If n is odd then $[n] = [2k+1] = [1]$. f is additive since the sum of two even or odd numbers is even and the sum of an even integer and an odd integer is odd, the product of two odd integers is odd and the product of an even integer with any integer is even.

The fiber of f above 0 is the set of even integers. The fiber of f above 1 is the set of odd integers.

(3) Fix $n \in \mathbb{Z}$. Is the map $f_n : \mathbb{Z} \to \mathbb{Z}$ defined by

$$f_n(x) = nx$$

a homomorphism? Why not? Consider $f(xy) = nxy$ and $f(x)f(y) = nxny = n^2xy$. (Thus f_n is a homomorphism if $n = 0, 1$, and f_0 is the zero morphism, f_1 is the identity)

(4) The map $g : \mathbb{R} \to M_2(\mathbb{R})$ given by

$$g(r) = \begin{pmatrix} 0 & 0 \\ -r & r \end{pmatrix}$$

Let's check this is a homomorphism. For any $r, s \in \mathbb{R}$ we compute

$$g(r) + g(s) = \begin{pmatrix} 0 & 0 \\ -r & r \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ -s & s \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ -r-s & r+s \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ -(r+s) & r+s \end{pmatrix} = g(r+s)$$

and

$$g(r)g(s) = \begin{pmatrix} 0 & 0 \\ -r & r \end{pmatrix}\begin{pmatrix} 0 & 0 \\ -s & s \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ -rs & rs \end{pmatrix} = g(rs)$$

What are the properties of g? Is it injective (Why)? Surjective (Why not)?

(5) Consider $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = x + 2$$

Is f a homomorphism? We have

$$f(a) + f(b) = (a+2) + (b+2) = (a+b+2) + 2 = f(a+b) + 2$$

or

$$f(a)f(b) = (a+2)(b+2) = ab + 2a + 2b + 4 = (ab+2) + 2a + 2b + 2 = f(ab) + 2a + 2b + 2$$

Theorem 3.31. *Let $f : R \to S$ be a homomorphism of rings. Then*

(1) $f(0_R) = 0_S$
(2) $f(-a) = -f(a)$ *for every* $a \in R$
(3) $f(a - b) = f(a) - f(b)$ *for all* $a, b \in R$

Proof. (1) $f(0_R) + f(0_R) = f(0_R + 0_R) = f(0_R) = f(0_R) + 0_S$ Thus

$$f(0_R) + f(0_R) = f(0_R) + 0_S$$

so from cancellation we have $f(0_R) = 0_S$.

(2) $f(a) + f(-a) = f(a - a) = f(0_R) = 0_S$. Thus $f(-a)$ is the solution to the equation $f(a) + x = 0_S$. We know that this solution (which is $-f(a)$) is unique. Therefore $f(-a) = -f(a)$.

(3) $f(a - b) = f(a + (-b)) = f(a) + f(-b) = f(a) + (-f(b)) = f(a) - f(b)$

□

(1) above says homomorphisms respect additive identities. However, they don't have to respect multiplicative identities! Give a simple example of a ring homomorphism $f : R \to S$ with doesn't map $1_R \to 1_S$. How about $f : \mathbb{R} \to M_2(\mathbb{R})$ given by

$$a \mapsto \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}$$

Then 1 does not map to $1_{M_2(\mathbb{R})}$. On the other hand,

$$f(1) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

is a unit for $f(\mathbb{R})$. This leads us to our next theorem.

Theorem 3.32. *If R is a ring with identity and f is surjective, then*

(1) *S is a ring with identity* $f(1_R)$
(2) *Whenever u is a unit in R then $f(u)$ is a unit in S and* $f(u)^{-1} = f(u^{-1})$

Proof. (1) We show that $f(1_R) \in S$ is the identity element. Let $x \in S$. Since f is surjective there exists $a \in R$ with $f(a) = x$. Then

$$f(1_R)x = f(1_R)f(a) = f(1_Ra) = f(a) = x$$

and

$$xf(1_R) = f(a)f(1_R) = f(a1_R) = f(a) = x$$

(2) Let u be a unit in R. We show that $f(u)$ is a unit in S. We compute

$$f(u)f(u^{-1}) = f(uu^{-1}) = f(1_R) = 1_S$$

and

$$f(u^{-1})f(u) = f(u^{-1}u) = f(1_R) = 1_S$$

□

3.4.4. *Ring Concepts.* Let $f : R \to S$ be a function. The *image* of f is defined as

$$\operatorname{Im} f = \{s \in S : s = f(r) \text{ for some } r \in R\} = \{f(r) : r \in R\}$$

Proposition 3.33. *Let $f : R \to S$ be a homomorphism of rings. Then the image of f (denoted $\operatorname{Im} f$) is a subring of S.*

Proof. $\operatorname{Im} f$ is nonempty since $0_S = f(0_R)$. Therefore we will invoke Theorem 3.6 in BH. Let $a, b \in \operatorname{Im} f$. Then $a = f(x), b = f(y)$ for $x, y \in R$. We have to show two things.

(1) Closure under subtraction. $a - b = f(x) - f(y) = f(x - y) \in \operatorname{Im} f$.
(2) Closure under multiplication. $ab = f(x)f(y) = f(xy) \in \operatorname{Im} f$.

□

Definition 3.34. The *kernel* of the ring homomorphism $f : R \to S$, denoted $\ker f$ is the set of elements that maps to zero, i.e.

$$\ker f = \{a \in R : f(a) = 0_S\}$$

Remark 3.35. *The kernel of a homomorphism measures the degree to which the homomorphism fails to be injective. This fact is captured by the next result, Proposition 3.36.*

Proposition 3.36. *Let $f : R \to S$ be a ring homomorphism. Then $\ker f = \{0_R\}$ if and only if f is injective.*

Proof. We first show ($\Longrightarrow$). Let $a, b \in R$ such that $f(a) = f(b)$. We wish to show that $a = b$. Since f is a homomorphism

$$\begin{aligned} f(a) - f(b) &= 0 \\ f(a - b) &= 0 \end{aligned}$$

Thus $a - b \in 0_R$.

We now show ($\Longleftarrow$). Let f be injective. We wish to show that $\ker f = \{0_R\}$. Let $a \in \ker f$. We have

$$f(a) = 0_S = f(0)$$

Then $a = 0$ since f is injective.

□

3.4.5. *Product Rings.* Let R, S be rings. Consider the cartesian product $R \times S$. This is a set, which we'll equip with addition and multiplication.

Proposition 3.37. *Consider* $R \times S$. *Define addition and multiplication component-wise*

$$(r, s) + (r', s') = (r + r', s + s') \qquad (r, s)(r', s') = (rr', ss')$$

Then $R \times S$ *is a ring. If* R, S *are both commutative then so is* $R \times S$. *If both* R, S *have an identity, then so does* $R \times S$.

For example, let's take $\mathbb{Z}_2 \times \mathbb{Z}_2$. We can look at the addition and multiplication tables below:

$$\begin{array}{c c} + & \begin{array}{cccc} (0,0) & (0,1) & (1,0) & (1,1) \end{array} \\ \begin{array}{c} (0,0) \\ (0,1) \\ (1,0) \\ (1,1) \end{array} & \begin{pmatrix} (0,0) & (0,1) & (1,0) & (1,1) \\ (0,1) & (0,0) & (1,1) & (1,0) \\ (1,0) & (1,1) & (0,0) & (0,1) \\ (1,1) & (1,0) & (0,1) & (0,0) \end{pmatrix} \end{array} \qquad \begin{array}{c c} \cdot & \begin{array}{cccc} (0,0) & (0,1) & (1,0) & (1,1) \end{array} \\ \begin{array}{c} (0,0) \\ (0,1) \\ (1,0) \\ (1,1) \end{array} & \begin{pmatrix} (0,0) & (0,0) & (0,0) & (0,0) \\ (0,0) & (0,1) & (0,0) & (0,1) \\ (0,0) & (0,0) & (1,0) & (1,0) \\ (0,0) & (0,1) & (1,0) & (1,1) \end{pmatrix} \end{array}$$

3.4.6. *Existence of Isomorphisms.* In general, there are no rules (algorithms) for determining whether two rings are isomorphic. One must construct an isomorphism.

Often it is equally important (and much easier) to demonstrate that two rings are *not* isomorphic. To do this, one has to show that there does not exist any isomorphism from one to the other. For this it is often useful to study invariants, or *properties preserved under isomorphism*. Often times that means that if $f : R \to S$ is an isomorphism and R obeys some property, then $f(R) = S$ obeys the same property. Sometimes this means that if $a \in R$ obeys a property then $f(a)$ obeys that property for isomorphism $f : R \to S$.

(1) Is $\mathbb{Z}_6$ isomorphic to $\mathbb{Z}_{12}$? To $\mathbb{Z}$? (Why not?) Easiest way is to compare cardinalities.

(2) Are $\mathbb{Z}_4$ and $\mathbb{Z}_2 \times \mathbb{Z}_2$ isomorphic? Their cardinalities are the same. Suppose so. Then there exists isomorphism $f : \mathbb{Z}_4 \to \mathbb{Z}_2 \times \mathbb{Z}_2$. Then $f(1) = (1, 1)$ (why ? since $1_{\mathbb{Z}_4} \mapsto 1_{\mathbb{Z}_2 \times \mathbb{Z}_2}$). Therefore

$$f(2) = f(1 + 1) = f(1) + f(1) = (1, 1) + (1, 1) = (2, 2) = (0, 0)$$

(3) In $\mathbb{Z}_8$ the elements $1, 3, 5, 7$ are units. Suppose $f : \mathbb{Z}_8 \to S$ is an isomorphism. Where do $f(1), f(3), f(5), f(7)$ get sent? f must map these four units to four units in S. In particular $\mathbb{Z}_8$ is not isomorphic to $\mathbb{Z}_4 \times \mathbb{Z}_2$ since this latter ring has only two units $(1, 1)$ and $(3, 1)$.

(4) Let R be a commutative ring and $f : R \to S$ be an isomorphism. Then for any $a, b \in R$ we can write $a = f(x), b = f(y)$ for $x, y \in R$ and

$$ab = f(x)f(y) = f(xy) = f(yx) = f(y)f(x) = ba$$

So S is commutative.

4. Polynomials and the ring $R[x]$

Bitte vergiß alles, was Du auf der Schule gelernt hast; denn Du hast es nicht gelernt.

– Edmund Landau[13]

The starting point for this chapter are the polynomials with 'coefficients' from a ring R. Just as Z_n was a model, the ring of polynomials $R[x]$ be also be a model for rings. As we cover more concepts in ring theory, apply them to $\mathbb{Z}_n$ and $R[x]$ to get an intuition for how they work.

4.1. **Polynomial Arithmetic.** We want to define a 'polynomial' in a rigorous fashion. Let R be a ring. A *polynomial with coefficients in* R is an expression of the form

$$a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n \tag{7}$$

where $n \geqslant 0$ (n nonnegative integer) and $a_i \in R$.

There are obvious questions: what is x? Is $x \in R$? Notice that the expression (7) makes sense provided that a_i and x all lie in some larger ring T that contains R, i.e. $R \subset T$.

For instance, $\pi \notin \mathbb{Z}$. But the expression $3 - 4\pi + 12\pi^2 + \pi^3$ makes sense in $\mathbb{R}$. Furthermore, it is not difficult to verify that the set of all numbers of the form

$$a_0 + a_1\pi + a_2\pi^2 + \ldots + a_n\pi^n \qquad n \geqslant 0, a_i \in \mathbb{Z}$$

is a subring of $\mathbb{R}$ that contains $\mathbb{Z}$ and π.

The next theorem is in the vein of this example. It lets us speak of polynomials rigorously.

Theorem 4.1. *Let R be a ring with identity. Then there exists a ring T containing an element $x \notin R$ such that:*

(1) *R is a subring of T*

(2) *$xa = ax$ for every $a \in R$*

(3) *The set $R[x]$ of all elements of T of the form*

$$a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n \qquad n \geqslant 0, a_i \in R$$

is a subring of T that contains R

(4) *The representation of elements of $R[x]$ is unique: if $n \geqslant m$ and*

$$a_0 + a_1 x_2 + a_2 x^2 + \ldots a_n x^n = b_0 + b_1 x + b_2 x^2 + \ldots + b_m x^m$$

then $a_i = b_i$ for $0 \leqslant i \leqslant n$ and $b_i = 0_R$ for $i > n$

(5) *$a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n = 0_R$ if and only if $a_i = 0_R$ for all i*

Proof. See BH Appendix G. Or see Hungerford (GTM) Chapter 3, Section 5. To give a flavor, consider the set of infinite sequences

$$(a_0, a_1, a_2, \ldots)$$

[13] Roughly: 'Please forget everything that you learned at school, because you have not learned it.' *Landau, E. (1965). Grundlagen Der Analysis: With Complete German-English Vocabulary (Vol. 141). American Mathematical Soc..*

such that $a_i \in R$ and only finitely many of the a_i are nonzero (i.e. there is some k such that $a_i = 0_R$ for all $i > k$)

The special term x will then be the element

$$(0, 1_R, 0, 0, \dots)$$

We think of

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

The left hand side is an infinite sequence. The right hand side is a polynomial. From this, we can see that the coefficient a_i of x^i holds the ith term in the sequence

We define addition component-wise:

$$(a_0, a_1, a_2, \dots) + (b_0, b_1, b_2, \dots) = (a_0 + b_0, a_1 + b_1, a_2 + b_2, \dots)$$

Multiplication is defined as follows:

$$(a_0, a_1, a_2, \dots)(b_0, b_1, b_2, \dots) = (c_0, c_1, c_2, \dots)$$

where $c_n = \sum_{i=0}^{n} a_{n-i} b_i$

□

The elements of the ring $R[x]$ are called *polynomials with coefficients in* R and the a_i are the *coefficients*. The element x is sometimes called an *indeterminate* or *formal variable*.

Remark 4.2. *Theorem 4.1 as stated needs R to be a ring with identity. This is required in order to contain the element $x = (0, 1_R, 0, \dots)$. However, we can also form polynomial rings that don't contain x, as we'll see below.*

Remark 4.3. *Property (2) does not imply that T is commutative.*

Observe that Property (5) is just a special case of (4), i.e. if $a_0 + a_1 x^1 + \dots + a_n x^n = 0_R x^0 + 0_R x^1 + \dots 0_R x^n$ then $a_i = 0_R$.

The expression in (5) is **not** *an equation to be 'solved' for x. x is not a variable, it is a specific element of a ring.*

4.1.1. *Basic Examples.* The rings $\mathbb{Z}[x], \mathbb{Q}[x], \mathbb{R}[x]$ are the rings of polynomials with which you are likely familiar.

Let E be the ring of even integers (sometimes denoted $2\mathbb{Z}$). Then $2 - 4x + 2x^2 \in E[x]$. However $x \notin E[x]$.

4.1.2. *Polynomial Arithmetic.* Addition for polynomials follow from the fact that $R[x]$ is a ring (which was given by Theorem 4.1).

Let's consider a more exotic example. Consider the ring $\mathbb{Z}_7[x]$. We'll do some addition in $\mathbb{Z}_7[x]$. Define $f(x) = 1 + 5x - x^2 + 4x^3 + 2x^4$ and $g(x) = 4 + 2x + 3x^2 + x^3$ in $\mathbb{Z}_7[x]$. Then

$$\begin{aligned} f(x) + g(x) &= (1 + 5x - x^2 + 4x^3 + 2x^4) + (4 + 2x + 3x^2 + x^3 + 0x^4) \\ &= (1+4) + (5+2)x + (-1+3)x^2 + (4+1)x^3 + (2+0)x^4 \\ &= 5 + 0x + 2x^2 + 5x^3 + 2x^4 \\ &= 5 + 2x^2 + 5x^3 + 2x^4 \end{aligned}$$

To do this computation we use commutativity, associativity and distributivity of the ring $R[x]$ (these properties follow from Theorem 4.1).

Let's consider multiplication in $\mathbb{Q}[x]$. Then

$$\begin{aligned}(1-7x+x^2)(2+3x) &= 1(2+3x)-7x(2+3x)+x^2(2+3x)\\ &= 1(2)+1(3x)-7x(2)-7x(3x)+x^2(2)+x^2(3x)\\ &= 2-11x-19x^2+3x^2\end{aligned}$$

This multiplication is accomplished by repeated application of distributivity.

These examples generalize. You add polynomials by adding the corresponding coefficients. Polynomial addition is given by the rule:

$$(a_0+a_1x+\ldots+a_nx^n)+(b_0+b_1x+\ldots+b_nx^n)=(a_0+b_0)+(a_1+b_1)x+\ldots+(a_n+b_n)x^n \tag{8}$$

Polynomial multiplication uses the distributive laws and collecting like powers of x. Polynomial multiplication is given by the rule:

$$(a_0+a_1x+a_2x^2+\ldots+a_nx^n)(b_0+b_1x+b_2x^2+\ldots b_mx^m) \tag{9}$$

$$=a_0b_0+(a_0b_1+a_1b_0)x+(a_0b_2+a_1b_1+a_2b_0)x^2+\ldots+a_nb_mx^{n+m} \tag{10}$$

For each $k \geqslant 0$, the coefficient of x^k in the product is given by the formula

$$a_0b_k+a_1b_{k-1}+\ldots+a_{k-1}b_1+a_kb_0=\sum_{i=0}^{k}a_ib_{k-i}$$

$a_i = 0_R$ if $i > n$ and $b_j = 0_R$ if $j > m$.

It is straightforward from this description of multiplication in $R[x]$ that if R is commutative then so is $R[x]$. Why?

Definition 4.4. Let $f(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n \in R[x]$ with $a_n \neq 0_R$. Then a_n is called the *leading coefficient* of $f(x)$. The *degree* of $f(x)$ is the integer n, denoted $\deg f(x)$ or $\deg f$, i.e. $\deg f$ is the largest exponent of x that appears with a nonzero coefficient, and this coefficient is the leading coefficient.

Let $f(x) = 3 - x + 4x^2 - 7x^3 \in \mathbb{R}[x]$. What is $\deg f$? $\deg f = 3$. What is the leading coefficient? -7.

What is $\deg(3 + 5x)$? $\deg(3 + 5x) = 1$. $\deg(x^{12}) = 12$.[14]

Remark 4.5. *The ring R (the coefficients) is a subring of the polynomial ring $R[x]$. The elements of R can be considered as polynomials in $R[x]$ using the map $R \to R[x]$ given by*

$$R \ni a \mapsto ax^0 \in R[x]$$

In fact, the map $a \mapsto ax^0$ is a monomorphism of rings (why?).

The polynomials of degree 0 in $R[x]$ are precisely the nonzero constant polynomials. The constant polynomial 0_R does not have a degree (no power of x appears with nonzero coefficient).

[14] One way to try to think about degree is as a function. However, as mentioned in class since $\deg(0_R)$ is undefined since no power of x appears with nonzero coefficient. To try to set up degree as a function it would be of the form $\deg : R[x]\backslash\{0_R\} \to \mathbb{N}$ where $R[x]\backslash\{0_R\}$ is the set of polynomials with coefficients in R with 0_R removed.

Remark 4.6. *There is some ambiguity in the use of notation here (and you will find this ambiguity present in BH):* a *may refer to either* $a \in R$ *or the constant polynomial* $a = ax^0 \in R[x]$*. This happens quite often with* 0_R *- this may refer to the zero polynomial in* $R[x]$ *or the zero element in* R*. From the context, it should always be clear which is meant.*

The following theorem will be our workhorse for this section!

Theorem 4.7. *If* R *is an integral domain with* $f(x), g(x)$ *nonzero polynomials in* $R[x]$*, then*

$$\deg[f(x)g(x)] = \deg f(x) + \deg g(x)$$

Proof. Assume $f(x) = a_0 + a_1x + a_2x^2 + \ldots a_nx^n$ and $g(x) = b_0 + b_1x + b_2x^2 + \ldots + b_mx^m$ with $a_n \neq 0_R$ and $b_m \neq 0_R$, so that $\deg f = n$ and $\deg g = m$. Then

$$f(x)g(x) = a_0b_0 + (a_0b_1 + a_1b_0)x + \ldots + a_nb_mx^{n+m}$$

The largest exponent of x that may have a nonzero coefficient is $n+m$. We have $a_nb_m \neq 0_R$ since R is an integral domain and $a_n \neq 0_R$ and $b_m \neq 0_R$. Therefore $f(x)g(x)$ is nonzero (therefore degree is defined) and $\deg[f(x)g(x)] = n + m = \deg f(x) + \deg g(x)$. □

We now list many corollaries that this theorem will give us.

Corollary 4.8. *If* R *is an integral domain, then so is* $R[x]$*.*

Proof. Since R is commutative we argued that $R[x]$ must be commutative. Why must $R[x]$ have an identity? Is 1_R an identity?

The proof of Theorem 4.7 shows that the product of nonzero polynomials in $R[x]$ is nonzero. Therefore, $R[x]$ is an integral domain. □

We also saw from the proof of Theorem 4.7 that the following holds:

Corollary 4.9. *Let* R *be a ring. If* $f(x), g(x)$ *and* $f(x)g(x)$ *are nonzero in* $R[x]$*, then*

$$\deg[f(x)g(x)] \leqslant \deg f(x) + \deg g(x)$$

Proof. The following equality (where f, g are defined above) holds in any ring

$$f(x)g(x) = a_0b_0 + (a_0b_1 + a_1b_0)x + \ldots + a_nb_mx^{n+m}$$

□

Let's work some examples. Consider $\mathbb{Z}_6[x]$. (Is $\mathbb{Z}_6$ an integral domain?) Let $f(x) = 2x^4$ and $g(x) = 5x$. Then $f(x)g(x) = (2x^4)(5x) = 4x^5$. Thus $\deg[fg] = \deg f + \deg g$.

However, if $g(x) = 1 + 3x^2$ then

$$f(x)g(x) = 2x^4(1 + 3x^2) = 2x^4 + 2(3x^6) = 2x^4 + 0x^6 = 2x^4$$

which has degree 4. An easy computation shows that $\deg f + \deg g = 6$. Thus the inequality below is strict

$$\deg fg < \deg f + \deg g$$

Corollary 4.10. *Let* R *be an integral domain and* $f(x) \in R[x]$*. Then* $f(x)$ *is a unit in* $R[x]$ *if and only if* $f(x)$ *is a constant polynomial that is a unit in* R*. In particular, if* F *is a field, the units in* $F[x]$ *are the nonzero constants in* F*.*

Proof. First assume that $f(x)$ is a unit in $R[x]$. Then $f(x)g(x) = 1_R$ for some $g(x) = 1_R$. Since R is integral domain, by Theorem 4.7 we have

$$\deg f + \deg g = \deg[fg] = \deg 1_R = 0$$

Since the degrees are nonnegative integers, we must have that $\deg f = 0$ and $\deg g = 0$. Therefore $f(x), g(x)$ are constant polynomials, i.e. elements of R. Since $f(x)g(x) = 1_R$ we have that $f(x)$ is a unit in R.

Now assume that $f(x)$ is a constant polynomial that is a unit in R, i.e. $f(x) = b$ with $b \in R$ a unit. Let $h(x) = b^{-1}$. Then $f(x)h(x) = bb^{-1} = 1_R = b^{-1}b = h(x)f(x)$. Therefore $f(x)$ is a unit in $R[x]$.

The last statement follows as every nonzero element of $a \in F$ is a unit when F is a field. Thus $a(x)$ is a unit.

□

Let's apply the Corollary 4.10. What do we know about the units in $\mathbb{Z}[x]$? The only units in $\mathbb{Z}$ are ± 1, thus only units in $\mathbb{Z}[x]$ are $1, -1$. The units in $\mathbb{R}[x], \mathbb{Q}[x], \mathbb{C}[x]$ are all nonzero constants since $\mathbb{R}, \mathbb{Q}, \mathbb{C}$ are fields.

As we've seen, polynomials over $\mathbb{Z}_n$ are more complicated. Consider $f(x) = 5x + 1$ in $\mathbb{Z}_{25}[x]$. $(5x+1)(20x+1) = 100x^2 + 25x + 1 = 0x^2 + 0x + 1 = 1$.

$x^2 + 1$ is not a perfect square in $\mathbb{Z}[x]$. However, it *is* a perfect square in $\mathbb{Z}_2[x]$, as in $\mathbb{Z}_2[x]$ we can compute

$$(x+1)(x+1) = x^2 + 2x + 1 = x^2 + 1$$

4.1.3. *Evaluation homomorphism.* There is an important homomorphism for polynomial rings. Let $\varphi : R[x] \to R$ be the map defined by

$$R[x] \ni f(x) \mapsto f(0) \in R$$

This is called the *evaluation map*.

Let's consider a specific example. Consider the evaluation map $\varphi : \mathbb{Q}[x] \to \mathbb{Q}$ defined by $\varphi(f(x)) = f(0)$. Let $f(x) \in \mathbb{Q}[x]$ with $f(x) = a_0 + a_1x^1 + \ldots + a_nx^n$. $\varphi(f(x)) = f(0) = a_0$. Therefore φ maps any polynomial $f(x)$ to its constant term. It's easy to see that this is a homomorphism since the constant term of the sum of two polynomials is the sum of their constant terms and the constant terms of the product two polynomials is the product of their constant terms (we verified this in Eqn. 8,9). The fiber above any $a \in \mathbb{Q}$ is the set of polynomials with a as their constant term. $\ker \varphi$ consists of the polynomials with constant term 0.

Remark 4.11. *The evaluation map appears in problem 18 in 4.1 of BH.*

Remark 4.12. *The evaluation map may be thought of as a projection from a polynomial onto its constant term. Consider the map $\phi : R[x] \to R$ given by $a_0 + a_1x + \ldots + a_nx^n \mapsto a_1$ i.e. a projection onto the coefficient of x^1. Is ϕ a homomorphism?*

4.2. Division Algorithm in $F[x]$.

In this section we consider $F[x]$ where F is some field.

Just as we had division in $\mathbb{Z}$ we also have division in $F[x]$.

Theorem 4.13. *Let F be a field and $f(x), g(x) \in F[x]$ with $g(x) \neq 0_F$. Then there exist unique polynomials $q(x)$ and $r(x)$ such that*

$$f(x) = g(x)q(x) + r(x) \qquad \textit{and either } r(x) = 0_F \qquad \textit{or} \qquad \deg r(x) < \deg g(x)$$

Proof. We will skip this proof. However it may be found as Theorem 4.6 in BH page 91. □

Can you restate the Division Algorithm for $\mathbb{Z}$ and draw the parallels between the two? We shall divide $f(x) = 3x^5 + 2x^4 + 2x^3 + 4x^2 + x - 2$ and $g(x) = 2x^3 + 1$.

Divisibility in $F[x]$

In this section F always denotes a field. We now port many of the results in Chapter 1.2 on divisibility in $\mathbb{Z}$.

Definition 4.14. Let F be a field and $a(x), b(x) \in F[x]$ with $b(x)$ nonzero. We say that $b(x)$ divides $a(x)$ and write $b(x)|a(x)$ if $a(x) = b(x)h(x)$ for some $h(x) \in F[x]$. We also say that $b(x)$ is a *factor* of $a(x)$.

Let's consider an example. Consider $\mathbb{Q}[x]$. Let $f(x) = 6x^2 - x - 2$. Then $(2x + 1)|f(x)$ in $\mathbb{Q}[x]$ since $(2x + 1)(3x - 2) = 6x^2 - x - 2$.

Notice that every constant multiple of $2x + 1$ also divides $f(x)$. For instance $5(2x + 1) = 10x + 5$ divides f as $5(2x + 1)\left[\frac{1}{5}(3x - 2)\right] = 6x^2 - x - 2$

This leads us to the next result.

Theorem 4.15. *Let F be a field and $a(x), b(x) \in F[x]$ with $b(x)$ nonzero.*

(1) *If $b(x)$ divides $a(x)$, then $cb(x)$ divides $a(x)$ for each nonzero $c \in F$*
(2) *Every divisor of $a(x)$ has degree less than or equal to $\deg a(x)$.*

Proof. (1) If $b(x)|a(x)$, then $a(x) = b(x)h(x)$ for some $h(x) \in F[x]$. Thus

$$a(x) = 1_F b(x)h(x) = cc^{-1}b(x)h(x) = cb(x)[c^{-1}h(x)]$$

Therefore $cb(x)|a(x)$.

(2) Suppose $b(x)|a(x)$. Then $a(x) = b(x)h(x)$. Thus

$$\deg a(x) = \deg b(x)h(x) \leqslant \deg b(x) + \deg h(x)$$

Since degrees are nonnegative we must have $0 \leqslant \deg b(x) \leqslant \deg a(x)$

□

We learned about gcd previously in $\mathbb{Z}$. We also have a notion of greatest common divisor of two polynomials $a(x), b(x) \in F[x]$. This should be a polynomial of highest degree that divides both of them. However, this is not necessarily unique, as any constant multiple is also a divisor by Theorem 4.15 Part (1).

We want the gcd to be unique (in fact we need it to be unique in order to speak about *the* greatest common divisor). One way to guarantee uniqueness is to require that a gcd be *monic*. $f(x) \in F[x]$ is *monic* if its leading coefficient is 1_F.

Definition 4.16. Let F be a field and $a(x), b(x) \in F[x]$, not both zero. The *greatest common divisor* of $a(x)$ and $b(x)$ is the monic polynomial of highest degree that divides $a(x)$ and $b(x)$. Formally, $d(x)$ is the gcd of $a(x), b(x)$ provided that $d(x)$ is monic and

(1) $d(x)|a(x)$ and $d(x)|b(x)$
(2) If $c(x)|a(x)$ and $c(x)|b(x)$ then $\deg c(x) \leqslant \deg d(x)$

Any two polynomials $a(x), b(x)$ have at least one monic common divisor, namely 1_F. Since the degree of a common divisor of $a(x)$ and $b(x)$ cannot exceed the $\deg a(x)$ or $\deg b(x)$ there must be *at least* one common monic (we will see from Bezout's theorem for polynomials below that there is in fact only one).

Let's do an example. Let $a(x) = 2x^4 + 5x^3 - 5x - 2$ and $b(x) = 2x^3 - 3x^2 - 2x$ in $\mathbb{Q}[x]$. We have the factorizations

$$\begin{aligned} a(x) &= (2x+1)(x+2)(x+1)(x-1) \\ b(x) &= (2x+1)(x-2)x \end{aligned}$$

$2x + 1$ is a common divisor, and of highest degree. Therefore in this case the monic $\frac{1}{2}(2x+1) = x + \frac{1}{2}$ is the gcd.

Theorem 4.17 (Bezout for Polynomials). *Let F be a field and $a(x), b(x) \in F[x]$, not both zero. Then there is a unique greatest common divisor $d(x)$ of $a(x)$ and $b(x)$. Furthermore, there are (not necessarily unique) polynomials $u(x)$ and $v(x)$ such that $d(x) = a(x)u(x) + b(x)v(x)$*

Proof. We'll skip this proof also. Note that it is appears as Theorem 4.8 in BH. □

Corollary 4.18. *Let F be a field and $a(x), b(x) \in F[x]$, not both zero. A monic polynomial $d(x) \in F[x]$ is the greatest common divisor of $a(x)$ and $b(x)$ if and only if $d(x)$ satisfies these conditions:*

(1) $d(x)|a(x)$ *and* $d(x)|b(x)$
(2) *if* $c(x)|a(x)$ *and* $c(x)|b(x)$ *then* $c(x)|d(x)$

Proof. Let $d(x)$ be the gcd. Then $d(x)|a(x), b(x)$ by definition. If $c(x)|a(x), b(x)$ then $a(x) = u(x)c(x)$ and $b(x) = v(x)c(x)$. By Bezout we have $d(x) = u'(x)a(x) + v'(x)b(x)$. Thus $d(x) = u'(x)u(x)c(x) + v'(x)v(x)c(x) = c(x)[u'(x)u(x) + v'(x)v(x)]$.

Now assume that $d(x)$ has the properties. We now wish to show $d(x)$ is gcd. We have $d(x)$ is monic and $d|a(x), b(x)$. Furthermore if $c(x)|a(x), b(x)$ then $d(x) = u(x)c(x)$. Thus $\deg d(x) = \deg u(x)c(x) = \deg u(x) + \deg c(x)$. Thus $\deg d(x) \geqslant \deg c(x)$.

□

Just as in the integers, $f(x)$ and $g(x)$ are said to be *relatively prime* if $\gcd(f, g) = 1_F$.

Proposition 4.19. *Let F be a field and $a(x), b(x) \in F[x]$. If $a(x)|b(x)c(x)$ and $a(x)$ and $b(x)$ are relatively prime, then $a(x)|c(x)$.*

Proof. By hypothesis $b(x)c(x) = k(x)a(x)$. Furthermore by Bezout we have

$$1_F = u(x)a(x) + v(x)b(x)$$

Multiply both sides by $c(x)$ to get

$$c(x) = c(x)[u(x)a(x) + v(x)b(x)] = c(x)u(x)a(x) + v(x)[b(x)c(x)] = c(x)u(x)a(x) + v(x)k(x)a(x)$$

Therefore we have

$$c(x) = a(x)[u(x)c(x) + v(x)k(x)]$$

□

5. Irreducibles and Unique Factorization

Let F be a field.

We're going to develop a notion of what it means for a polynomial $f(x) \in F[x]$ to be 'prime'. We're going to call this 'irreducible'.

Recall that a nonzero $p \in Z$ is prime if $p \neq \pm 1$ and the only divisors of p are ± 1 and $\pm p$, i.e. $p = (\pm p)(\pm 1)$

Here's a more general notion. Let R be a commutative ring with identity. $a \in R$ is said be an *associate* of $b \in R$ if $a = bu$ for some unit $u \in R$. In this case b is an associate of a since $b = au^{-1}$.

In the ring $\mathbb{Z}$ the only associates of an integer n are $n, -n$, since the only units are ± 1.

Now from Corollary 4.10 the units in $F[x]$ are the nonzero constants. Therefore $f(x)$ is an associate of $g(x)$ in $F[x]$ if and only if $f(x) = cg(x)$ for some nonzero $c \in F$.

Going back to the definition of prime, we have that p is prime if its only divisors are ± 1 (the units) and $\pm p$ (the associates). Therefore we'll make the following definition for $F[x]$:

Definition 5.1. Let F be a field. A nonconstant polynomial $p(x) \in F[x]$ is said to be *irreducible* if its only divisors are its associates and the nonzero constant polynomials (units). A nonconstant polynomial that is not irreducible is called *reducible*.

As mentioned in BH, we could just as well use the word 'prime' instead of 'irreducible'. 'Irreducible' is customary for polynomials. However, both of these words express the same concept: an 'atomistic' behavior. Irreducible/prime objects *cannot be expressed in simpler terms*.

Let's look at a concrete example. We show that polynomial $g(x) = x + 2$ is irreducible in $\mathbb{Q}[x]$. We know that since $\mathbb{Q}$ is a field (and in particular an integral domain) we have that divisors of $g(x)$ must have degree 0 or 1. If $f(x) | x + 2$ then $f(x)a(x) = g(x)$. If $\deg f(x) = 0$ then $f(x)$ is a nonzero constant. If $\deg f(x) = 1$ then $\deg g(x) = 0$, so $g(x) = c$. Thus $c^{-1}(x + 2) = f(x)$ and $f(x)$ is an associate of $g(x)$. Therefore g is irreducible. A general argument proceeds along similar lines, and shows that every polynomial of degree 1 in $F[x]$ is irreducible in $F[x]$.

Theorem 5.2. *Let F be a field. A nonzero polynomial $f(x)$ is reducible in $F[x]$ if and only if $f(x)$ can be written as the product of two polynomials of lower degree.*

Proof. We'll first do ($\Longrightarrow$). Assume that $f(x)$ is nonzero and reducible. Then $f(x)$ has a divisor $g(x)$ and $g(x)$ is not an associate and $g(x)$ is not a nonzero constant. We can write $f(x) = g(x)h(x)$. If $\deg g = \deg f$ then since $\deg g = \deg f = \deg gh = \deg + \deg h$. We must have that $\deg h = 0$. If $\deg h = \deg f$ then $\deg g = 0$ by a similar argument.

If $\deg g = 0$, then $g(x)$ is a nonzero constant. If $\deg h = 0$ then h is a nonzero constant and thus $g(x)$ is an associate. In either case our hypothesis is violated. Thus $\deg g < \deg f$ and $\deg h < \deg f$.

Now we'll do ($\Longleftarrow$). Assume that $f(x)$ can be written as the product of two polynomials of lower degree, i.e. $f(x) = g(x)h(x)$. We have the equality $\deg f(x) = \deg g(x) + \deg h(x)$. If $\deg g = 0$, this forces $\deg h = \deg f$, contradicting our hypothesis. Thus g cannot be a nonzero constant. If g is an associate, then $\deg g = \deg f$, another contradiction. There $g(x)$ is a divisor that is neither a nonzero constant nor an associate. □

Remark 5.3. *The concept of reducibility depends on the particular polynomial ring. Notice that in $\mathbb{C}[x]$ we have $x^2+1=(x+i)(x-i)$. Therefore x^2+1 is reducible. However in $\mathbb{Q}[x]$ one can show that x^2+1 is irreducible.*

Theorem 5.4. *Let F be a field and $p(x) \in F[x]$ be a nonconstant polynomial. Then the following are equivalent:*

(1) *$p(x)$ is irreducible*
(2) *If $b(x), c(x)$ are polynomials such that $p(x)|b(x)c(x)$ then $p(x)|b(x)$ or $p(x)|c(x)$*
(3) *If $r(x)$ and $s(x)$ are any polynomials such that $p(x)=r(x)s(x)$ then $r(x)$ or $s(x)$ is a nonzero constant polynomial*

Proof. (1) We show $(1) \implies (2)$. Let $d(x)=(p(x),b(x))$. Then $d(x)=1_F$ or $d(x)=cp(x)$ since $p(x)$ is irreducible. If $d(x)=cp(x)$ then $p(x)|b(x)$ by definition of gcd. If $d(x)=1_F$, then $p(x)|c(x)$ by Proposition 4.19.

(2) We show $(2) \implies (3)$. Assume that $p(x)=r(x)s(x)$. By our property we have that $p(x)|r(x)$ or $p(x)|s(x)$. If $p(x)|r(x)$ then we have $\deg p(x)=\deg r(x)+\deg s(x)$ and $\deg p(x) \leqslant \deg r(x)$. Thus $\deg r(x)+\deg s(x)=\deg p(x) \leqslant \deg r(x)$. This implies that $\deg s(x)=0$, i.e. $s(x)$ is a nonzero constant polynomial.

(3) We show that $p(x)$ is irreducible. Write $p(x)=r(x)s(x)$. Then $r(x)$ or $s(x)$ is a nonzero constant polynomial. Without loss of generality, assume that $r(x)=c \in F$. Then $r(x)$ is a unit. Thus $s(x)$ is an associate of $p(x)$. Therefore any factor is an associate or a unit, so $p(x)$ is irreducible.

□

Corollary 5.5. *Let F be a field and $p(x)$ an irreducible polynomial in $F[x]$. If $p(x)|a_1(x)a_2(x)\ldots a_n(x)$, then $p(x)$ divides at least one of the $a_i(x)$.*

Proof. We use Theorem 5.4. Since $p(x)$ is irreducible either $p(x)|a_1(x)$ or $p(x)|a_2(x)\ldots a_n(x)$. If $p(x) \not| a_1(x)$ then repeat the argument on $a_2(x)\ldots a_n(x)$.

□

Theorem 5.6. *Let F be a field. Every nonconstant polynomial $f(x)$ in $F[x]$ is the product of irreducible polynomials in $F[x]$. This factorization is unique in the following sense: if*

$$f(x)=p_1p_2(x)\cdots p_t(x) \qquad \textit{and} \qquad f(x)=q_1(x)q_2(x)\cdots q_s(x)$$

with each $p_i(x)$ and $q_j(x)$ irreducible, then $r=s$. Without loss of generality (meaning up to some relabeling) $p_i(x)$ is an associate of $q_i(x)$

Proof. We'll repeat basically the same argument as Theorem 1.7 in BH.

Let S be the set of all nonconstant polynomials that are not the product of irreducibles. Let $T=\{\deg f(x) : f(x) \in S\}$. We will use proof by contradiction to show that S is empty. Suppose that S is nonempty. Then T is nonempty. T consists of nonnegative integers (why?), thus must have a least element $n \in T$ by well-ordering principle. This must correspond to some $f(x) \in S$ with $\deg f(x)=n$. Since $f(x) \in S$, S is not irreducible, thus $f(x)=a(x)b(x)$ where $a(x), b(x)$ are neither units nor associates. Thus $0<\deg a(x), \deg b(x)<\deg f(x)$. This implies $a(x), b(x) \notin S$. Therefore they are the product of irreducibles, so $f(x)$ is the product of irreducibles. This is a contradiction.

Uniqueness follows along similar lines, as can be found in Theorem 4.14.

□

5.1. **Polynomial Functions, Roots, and Reducibility.** We have considered the parallels between $F[x]$ and $\mathbb{Z}$ - namely, irreducibles/primes, the division algorithm and factorizations.

We didn't cover primality testing in $\mathbb{Z}$, however we will do more in $F[x]$ as we have more structure.

In particular, notice that every polynomial in $F[x]$ induces a function from F to F. The properties of this function (in particular, its roots) are related to reducibility of the polynomial.

Associated with each polynomial $a_n x^n + \ldots + a_1 x + a_0 \in F[x]$ is a function $f : F \to F$ given by $f(r) = a_n r^n + \ldots + a_1 r + a_0$. This can be thought of as plugging in values of F into the polynomial.

Here are some examples. Consider the polynomial $x^2 + 5x + 3 \in \mathbb{R}[x]$. This polynomial induces a function $f : \mathbb{R} \to \mathbb{R}$ given by $f(r) = r^2 + 5r + 3$.

Consider the polynomial $x^4 + x + 1 \in \mathbb{Z}_3[x]$. This induces the function $f : \mathbb{Z}_3 \to \mathbb{Z}_3$ given by $f(r) = r^4 + r + 1$. Since $\mathbb{Z}_3$ is small, we can go ahead and evaluate all the values of f. Namely,

$$f(0) = 0^4 + 0 + 1 = 1, \quad f(1) = 1^4 + 1 + 1 = 0, \quad f(2) = 16 + 2 + 1 = 1$$

The polynomial $x^3 + x^2 + 1 \in \mathbb{Z}_3[x]$ induces the function $g : \mathbb{Z}_3 \to \mathbb{Z}_3$ given by

$$g(0) = 0 + 0 + 1, \quad g(1) = 1^3 + 1^2 + 1 = 0, \quad g(2) = 8 + 4 + 1 = 1$$

Notice that f, g are the same function induced on $\mathbb{Z}_3$, even though they are induced by different polynomials in $\mathbb{Z}_3[x]$.

Definition 5.7. *Let R be a commutative ring and $f(x) \in R[x]$. An element $a \in R$ is a* root *for the polynomial $f(x)$ if $f(a) = 0_R$, i.e. the induced function $f : R \to R$ maps a to 0_R.*

Example: the polynomial $x^2 + 1 \in \mathbb{R}[x]$ has no roots in $\mathbb{R}$. However, $x^2 + 1 \in \mathbb{C}[x]$ has roots in $\mathbb{C}$, namely $i, -i$.

Theorem 5.8 (The Remainder Theorem)**.** *Let F be a field. Let $f(x) \in F[x]$ and $a \in F$. The remainder when $f(x)$ is divided by the polynomial $x - a$ is the polynomial $f(a)$.*

Let's consider some example. To find the remainder when $f(x) = x^{79} + 3x^{24} + 5$ is divided by $x - 1$. We apply the remainder theorem with $a = 1$. We have

$$f(1) = 1^{79} + 3(1)^{24} + 5 = 1 + 3 + 5 = 9$$

How about the remainder when $f(x) = 3x^4 - 8x^2 + 11x + 1$ is divided by $x + 2$? (Apply the remainder theorem carefully, it strictly says $x - a$). Thus we put the divisor in the form $x - (-2)$ and compute $f(-2) = 48 - 32 - 22 + 1 = -5$

Proof. By the Division Algorithm for Polynomials, $f(x) = (x - a)q(x) + r(x)$ where $r(x) = 0_F$ or $\deg r(x) < 1$. Notice that if $\deg r(x) < 1$ then $\deg r = 0$. Therefore in either case $r(x)$ is a constant polynomial, i.e. $r(x) = c$ for some $c \in F$. Thus $f(x) = (x - a) + c$. We plug in a to get $f(a) = (a - a) + c$ implying $c = f(a)$.

□

Theorem 5.9 (The Factor Theorem)**.** *Let F be a field. Let $f(x) \in F[x]$ and $a \in F$. Then a is a root of the polynomial $f(x)$ if and only if $x - a$ is a factor of $f(x)$ in $F[x]$.*

Proof. We first assume that a is a root of $f(x)$. This means that $f(a) = 0_F$. Let's apply the Division algorithm:

$$\begin{aligned} f(x) &= (x-a)q(x) + r(x) \\ f(x) &= (x-a)q(x) + f(a) \text{ here we're using Theorem 5.8} \\ f(x) &= (x-a)q(x) + 0_F \text{ since } f(a) \text{ is a root} \end{aligned}$$

Thus $(x-a)|f(x)$.

Now assume that $x-a$ is a factor for $f(x)$. We want to show that a is a root. Since $x-a$ is a factor we can write $f(x) = (x-a)q(x)$. Plug in a we have $f(a) = (a-a)q(a) = 0_F$. Thus a is a root.

□

Here's an example application of the theorem. Let's show that

$$f(x) = x^4 - x^3 + 2x^2 + 2x - 4$$

is reducible in $\mathbb{Q}[x]$. Notice that 1 is a root since $f(1) = 1 - 1 + 2 + 2 - 4$. Invoke Theorem 5.9 to get that $x-1$ is a divisor.

This gives us a very simple test for reducibility: if you can find a root in F for $f(x) \in F[x]$ then $f(x)$ is reducible.

Corollary 5.10. *Let F be a field and $f(x)$ a nonzero polynomial of degree n in $F[x]$. Then $f(x)$ has at most n roots in F.*

Proof. If $f(x)$ has a root in F, then by the factor theorem we know $x-a$ is a divisor. Thus $f(x) = (x-a)g(x)$. We have that $\deg f = 1 + \deg g(x)$. Therefore $\deg g(x) = n-1$. We can now repeat the argument on $g(x)$. □

Corollary 5.11. *Let F be a field and $f(x) \in F[x]$ with $\deg f(x) \geqslant 2$. If $f(x)$ is irreducible in $F[x]$ then $f(x)$ has no roots in F.*

Proof. If $f(x)$ is irreducible. If $a \in F$ was a root, then $x-a$ would be a factor by Theorem 5.9. However, $f(x)$ is irreducible. Therefore $f(x)$ has no roots in F. □

The converse of this corollary is false. Why? The converse says 'if $f(x)$ has no roots then $f(x)$ is irreducible'. Take some product of irreducible polynomials: $(x^2+1)(x^2+1) = 4x^2 + 2x + 1$ is reducible in $\mathbb{Q}[x]$ but has no roots in $\mathbb{Q}$.

On the other hand, the converse is true for degrees $2, 3$.

Corollary 5.12. *Let F be a field and let $f(x) \in F[x]$ be a polynomial of degree 2 or 3. Then $f(x)$ is irreducible in $F[x]$ if and only if $f(x)$ has no roots in F.*

Proof. Assume that $f(x)$ is irreducible. Then as we showed, f has no roots in F.

Now assume that $f(x)$ has no roots in F. Suppose that f is reducible. Then we can write $f(x) = a(x)b(x)$ for $a(x), b(x) \in F[x]$ with $0 < \deg a(x), \deg b(x) \leqslant \deg f$. We have that $\deg f = \deg a(x)b(x) = \deg a(x) + \deg b(x)$. If $\deg f = 2$, then $\deg a(x) = \deg b(x) = 1$. This implies $a(x)$ is of the form $cx + d$ for some $c, d \in F$. Therefore $-c^{-1}d$ is a root. If $\deg f(x) = 3$ then without loss of generality $\deg a(x) = 1$ and $\deg b(x) = 2$. Therefore we again have that $a(x)$ has form $cx + d$ for $c, d \in F$, which has root $-c^{-1}d$. This is a contradiction, therefore f is irreducible.

□

Let's do another example. Consider $f(x) = x^3 + x + 1$ in $\mathbb{Z}_5[x]$. Is $f(x)$ irreducible? Since $\deg f(x) = 3$, we can apply Corollary 5.12. We compute

$$f(0) = 1, \quad f(1) = 3, \quad f(2) = 1, \quad f(3) = 1, \quad f(4) = 4$$

Therefore by Corollary 5.12 $f(x)$ is irreducible in $\mathbb{Z}_5[x]$.

Corollary 5.13. *Let F be an infinite field and $f(x), g(x) \in F[x]$. Then $f(x)$ and $g(x)$ induce the same function from F to F if and only if $f(x) = g(x)$ in $F[x]$.*

Proof. Suppose that $f(x)$ and $g(x)$ induce the same function from F to F. Then $f(a) = g(a)$, so $f(a) - g(a) = 0_F$ for all $a \in F$. Therefore every element of F is a root of the polynomial $f(x) - g(x)$. By hypothesis F is infinite. However, if $f(x) - g(x)$ is nonzero then by Corollary 5.10 we know it can have at most $\deg f(x)$ roots. Therefore $f(x) - g(x) = 0_F$.

If $f(x) = g(x)$ in $F[x]$, then they induce the same function from F to F.

□

5. Congruence in $F[x]$

> *A good stock of examples, as large as possible, is indispensable for a thorough understanding of any concept, and when I want to learn something new, I make it my first job to build one.*
>
> – Paul Halmos[15]

The ideas of congruence in $\mathbb{Z}$ that we explored in Chapter 2 carry over to the polynomial ring $F[x]$. In this section we continue building a 'stock of examples' for the ideas of congruence and quotients. In the next chapter we will abstract these to general rings.

5.1. **Congruence.**

We've already seen that $F[x]$ obeys the same divisibility properties that $\mathbb{Z}$ had. It is not surprising then that many of the ideas of congruence can be ported directly.

Definition 5.1. *Let F be a field and $f(x), g(x), p(x) \in F[x]$ with $p(x)$ nonzero. Then $f(x)$* **is congruence to** $g(x)$ **modulo** $p(x)$ *written $f(x) \equiv g(x) \mod p(x)$ provided that $p(x)$ divides $f(x) - g(x)$.*

Remark 5.2. *Compare this definition with the definition of congruence class in $\mathbb{Z}_n$!*

Here's an example. Consider $\mathbb{Q}[x]$.

$$x^2 + x + 1 \equiv x + 2 \mod x + 1$$

since

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

Theorem 5.3. *Let F be a field and $p(x)$ be a nonzero polynomial in $F[x]$. Then the relation of congruence modulo $p(x)$ is:*

(1) *Reflexive:* $f(x) \equiv f(x) \mod p(x)$.
(2) *Symmetric: if* $f(x) \equiv g(x) \mod p(x)$ *then* $g(x) \equiv f(x) \mod p(x)$
(3) *Transitive: if* $f(x) \equiv g(x) \mod p(x)$ *and* $g(x) \equiv h(x) \mod p(x)$ *then* $f \equiv h(x) \mod p(x)$

Proof. (1) $f(x) - f(x) = 0_F p(x)$ Thus $f(x) \equiv f(x) \mod p(x)$
(2) If $f(x) - g(x) = q(x)p(x)$. Then $g(x) - f(x) = -(f(x) - g(x)) = -(q(x)p(x)) = (-q(x))p(x)$
(3) If $f(x) - g(x) = q(x)p(x)$ and $g(x) - h(x) = r(x)p(x)$ then

$$f(x) - h(x) = f(x) - g(x) + g(x) - h(x) = q(x)p(x) + r(x)p(x) = (q(x) + r(x))p(x)$$

□

Theorem 5.4. *Let F be a field and $p(x) \in F[x]$ a nonzero polynomial. If $f(x) \equiv g(x) \mod p(x)$ and $h(x) \equiv k(x) \mod p(x)$ then*

(1) $f(x) + h(x) \equiv g(x) + k(x) \mod p(x)$
(2) $f(x)h(x) \equiv g(x)k(x) \mod p(x)$

This proof follows the proof of Theorem 2.2 in BH.

Proof. By hypothesis we have $f(x) - g(x) = a(x)p(x)$ and $h(x) - k(x) = b(x)p(x)$.

[15] Quoted in *Gallian, J. (2016). Contemporary abstract algebra. Cengage Learning.*

(1) We compute

$$f(x) - h(x) - [g(x) - k(x)] = [f(x) - g(x)] - [h(x) - k(x)] = a(x)p(x) - b(x)p(x) = [a(x) - b(x)]p(x)$$

(2) We compute

$$\begin{aligned} f(x)h(x) - g(x)k(x) &= f(x)h(x) - f(x)k(x) + f(x)k(x) - g(x)k(x) \\ &= f(x)[h(x) - k(x)] + [f(x) - g(x)]k(x) \\ &= f(x)b(x)p(x) + a(x)p(x)k(x) = [f(x)b(x) + a(x)p(x)]k(x) \end{aligned}$$

□

Remark 5.5. *Again, notice the familiar trick of adding zero. This is very prominent in algebra.*

Definition 5.6. *Let F be a field and $f(x), p(x) \in F[x]$ with $p(x)$ nonzero. The* **congruence class** *of $f(x)$ modulo $p(x)$ is denoted $[f(x)]$ and consists of all polynomials in $F[x]$ that are congruent to $f(x)$ modulo $p(x)$, that is*

$$[f(x)] = \{g(x) : g(x) \in F[x] \text{ and } g(x) \equiv f(x) \mod p(x)\}$$

We can characterize $[f(x)]$ with a quick computation. Since $g(x) \equiv f(x) \mod p(x)$ means $g(x) - f(x) = k(x)p(x)$ we have $g(x) = f(x) + k(x)p(x)$ thus

$$\begin{aligned} [f(x)] &= \{g(x) : g(x) \equiv f(x) \mod p(x)\} \\ &= \{f(x) + k(x)p(x) : k(x) \in F[x]\} \end{aligned}$$

Remark 5.7. *Compare this congruence class of polynomials to the congruence class of an integer $[a]_n = \{a + kn : k \in \mathbb{Z}\}$!*

Let's work an example. Consider congruence mod $x^2 + 1$ in $\mathbb{R}[x]$. The congruence class of $2x + 1$ is the set

$$[2x + 1] = \{(2x + 1) + k(x)(x^2 + 1) : k(x) \in \mathbb{R}[x]\}$$

Theorem 5.8. $f(x) \equiv g(x) \mod p(x)$ *if and only if* $[f(x)] = [g(x)]$

Proof. We first show ($\implies$). Let $f(x) - g(x) = a(x)p(x)$. Let $h(x) \in [f(x)]$. We want to show $h(x) \in [g(x)]$. We have $h(x) = f(x) + k(x)p(x)$. Then $h(x) = g(x) + a(x)p(x) + k(x)p(x) = g(x) + [a(x) + k(x)]p(x)$. Thus $h(x) \in [g(x)]$. Since congruence modulo p is symmetric, we have $g(x) \equiv f(x) \mod p(x)$, and we can repeat the argument.

Now assume $[f(x)] = [g(x)]$. Since $f(x) \in [f(x)] = [g(x)]$ we have $f(x) = g(x) + k(x)p(x)$. Thus $f(x) - g(x) = k(x)p(x)$ and $f(x) \equiv g(x) \mod p(x)$.

□

Corollary 5.9. *Either $[f(x)] \cap [g(x)] = \emptyset$ or $[f(x)] = [g(x)]$.*

Proof. If $[f(x)] = [g(x)] = \emptyset$ then we're done. Otherwise there exists some $a(x) \in [f(x)] \cap [g(x)]$. We first show $[f(x)] \subset [g(x)]$. Let $h(x) \in [f(x)]$. We want to show $h(x) \in [g(x)]$. $h(x) \equiv f(x) \mod p(x)$, $f(x) \equiv a(x) \mod p(x)$ (by symmetry) and $a(x) \equiv g(x) \mod p(x)$. Therefore by transitivity we have $h(x) \equiv g(x) \mod p(x)$. We can repeat a similar argument for $[g(x)] \subset [f(x)]$. □

Recall that $\mathbb{Z}_n$ has precisely n elements, namely $[0], [1], \ldots, [n-1]$. There is a class for each remainder under division by n. We have the Division Algorithm in $F[x]$. In $F[x]$ the possible remainders under division by a polynomial of degree n are all the polynomials of degree less than n and 0_F. Therefore we have:

Corollary 5.10. *Let F be a field and $p(x) \in F[x]$ with $\deg p(x) = n \in \mathbb{Z}$*

(1) *If $f(x) \in F[x]$ and $r(x)$ is the remainder when $f(x)$ is divided by $p(x)$ then $[f(x)] = [r(x)]$*

(2) *Let*

$$S = \{f(x) \in F[x] : \deg f(x) < n\} \cup \{0_F\}$$

i.e. the set consisting of the zero polynomial and all the polynomials of degree less than n in $F[x]$. If $f(x) \in F[x]$ then $[f(x)] = [q(x)]$ for some $q(x) \in S$ and the congruence classes of different polynomials in S are distinct.

Proof. (1) By the division algorithm for polynomials, we have $f(x) = q(x)p(x) + r(x)$ where $r(x) = 0_F$ or $\deg r(x) < n$. Thus $f(x) - r(x) = q(x)p(x)$ and $f(x) \equiv r(x) \mod p(x)$. Therefore $[f(x)] = [r(x)]$ by Theorem 5.8.

(2) By part (1), $[f(x)] = [r(x)]$ and $r(x) = 0_F$ or $\deg r(x) < n$. Thus $r(x) \in S$. For $a(x), b(x) \in S$ we have $\deg[a(x) - b(x)] < n$, therefore $p(x)$ cannot divide $a(x) - b(x)$ and $a(x) \not\equiv b(x) \mod p(x)$. Thus they are distinct by Theorem 5.8.

□

Let's consider some more examples. Consider congruence modulo $x^2 + 1$ in $\mathbb{R}[x]$. As we've said, there is a congruence class for 0_F and for $f(x)$ where $\deg f < 2$. Therefore the congruence classes must be of the form $ax + b$ $a, b \in \mathbb{R}$ (possibly zero). Thus $\mathbb{R}[x]/(x^2 + 1)$ consists of infinitely many distinct congruence classes. Corollary 5.10 says that $[ax + b] = [cx + d]$ if and only if $ax + b = cx + d$. By the definition of equivalence in polynomials, this means that $a = c$ and $b = d$ (the coefficients are equal). Therefore every element of $R[x]/(x^2 + 1)$ can be uniquely written as $[ax + b]$.

Consider congruence modulo $x^2 + x + 1$ in $\mathbb{Z}_2[x]$. The possible remainders on division by $x^2 + x + 1$ are polynomials of the form $ax + b$ with $a, b \in \mathbb{Z}_2[x]$. Thus there are four polynomials:

$$0_F, \quad 1, \quad x, \quad x + 1$$

Thus $\mathbb{Z}_2[x]/(x^2 + x + 1)$ consists of four congruence classes $[0], [1], [x], [x + 1]$.

Congruence Class Arithmetic

Congruence in the integers led to the construction of $\mathbb{Z}_n$ and modular arithmetic. Similarly, congruence in $F[x]$ will produce new 'quotient rings'.

Theorem 5.11. *Let F be a field and $p(x)$ be a nonconstant polynomial in $F[x]$. If $[f(x)] = [g(x)]$ and $[h(x)] = [k(x)]$ in $F[x]/(p(x))$ then*

$$[f(x) + h(x)] = [g(x) + k(x)] \qquad \text{and} \qquad [f(x)h(x)] = [g(x)k(x)]$$

Proof. We have $f(x) \equiv g(x) \mod p(x)$ and $h(x) \equiv k(x) \mod p(x)$. Thus by Theorem 5.11 $f(x) + h(x) \equiv g(x) + k(x) \mod p(x)$ and $f(x)h(x) \equiv g(x)k(x) \mod p(x)$. Thus $[f(x) + h(x)] = [g(x) + k(x)]$ and $[f(x)h(x)] = [g(x)k(x)]$. □

We'll now define addition and multiplication of congruence classes for polynomials.

Definition 5.12. *Let F be a field and $p(x)$ be a nonconstant polynomial in $F[x]$. Addition and multiplication in $F[x]/(p(x))$ are defined by*

$$[f(x)] + [g(x)] = [f(x) + g(x)] \qquad\qquad [f(x)][g(x)] = [f(x)g(x)]$$

Let's again consider congruence x^2+1 in $\mathbb{R}[x]$.

The sum of the classes $[2x+1]$ and $[3x+5]$ is the class $[2x+1]+[3x+5]=[5x+6]$.

The product is $[2x+1][3x+5] = [(2x+1)(3x+5] = [6x^2+13x+5]$. We can divide $6x^2+13x+5$ by x^2+1 to get

$$\begin{aligned} 6x^2+13x+5 &= 6(x^2+1)+(13x-1) \\ 6x^2+13x+5-(13x-1) &= 6(x^2+1) \\ 6x^2+13x+5 &\equiv 13x-1 \quad \text{mod } x^2+1 \\ [6x^2+13x+5] &= [13x-1] \end{aligned}$$

Let's consider congruence in $\mathbb{Z}_2[x]/(x^2+x+1)$. We've seen that this has four congruence classes, namely $[0],[1],[x],[x+1]$. We can compute the addition and multiplication tables:

$$\begin{array}{c|cccc} + & [0] & [1] & [x] & [x+1] \\ \hline [0] & [0] & [1] & [x] & [x+1] \\ [1] & [1] & [0] & [x+1] & [x] \\ [x] & [x] & [x+1] & [0] & [1] \\ [x+1] & [x+1] & [x] & [1] & [0] \end{array} \qquad \begin{array}{c|cccc} \cdot & [0] & [1] & [x] & [x+1] \\ \hline [0] & [0] & [0] & [0] & [0] \\ [1] & [0] & [1] & [x] & [x+1] \\ [x] & [0] & [x] & [x^2]^{=[x+1]} & [x^2+x]^{=[1]} \\ [x+1] & [0] & [x+1] & [x^2+x]^{=[1]} & [x^2+2x+1]^{=[x]} \end{array}$$

Notice that $x^2+x = (x^2+x+1)+1$. Thus $x^2x+1 \equiv 1$. So $[x^2+x] = [1]$. We have $x^2 = (x^2+x+1)+(x+1)$ thus $[x][x] = [x^2] = [x+1]$. Furthermore $(x+1)(x+1) = x^2+2x+1 = x^2+1 = (x^2+x+1)+x$. Thus $[x+1][x+1] = [x^2+2x+1] = [x]$.

Let's look at the tables we've just computed. We can see that $\mathbb{Z}_2[x]/(x^2+x+1)$ is a commutative ring (What's an easy way to check this? Observe that the tables observe a symmetry across the diagonal). Furthermore it has a multiplicative identity $[1]$. Observe that $F^* = \{[0],[1]\}$ is a subring. What familiar ring does this look like? It is isomorphic to $\mathbb{Z}_2$. This is a general result, as we record in the next theorem.

Theorem 5.13. *Let F be a field and $p(x)$ be a nonconstant polynomial in $F[x]$. Then $F[x]/(p(x))$ is a commutative ring with identity. Moreover $F[x]/(p(x))$ contains a subring F^* that is isomorphic to F.*

Proof. It is fairly straightforward to work out the ring axioms. However, another way to do this is to write $F[x]/(p(x))$ as the homomorphic image of another ring. Consider the map $\varphi : F[x] \to F[x]/(p(x))$ given by

$$F[x] \ni f(x) \mapsto [f(x)] \in F[x]/(p(x))$$

We have that $F[x]/(p(x))$ is a set with two operations. If we verify that φ is a homomorphism then we have that $\text{Im}\varphi = \varphi(\mathbb{F}[x]) = F[x]/(p(x))$ is a ring (Why?).[16] We have this as a homomorphism since

$$\varphi(f(x)+g(x)) = [f(x)+g(x))] = [f(x)]+[g(x)] = \varphi(f(x))+\varphi(g(x))$$

[16] This is a slight abuse of definitions. Recall that a homomorphism was defined between two rings. However, it could as well have been defined with domain a ring and codomain a set equipped with addition and multiplication

and

$$\varphi(f(x)g(x)) = [f(x)g(x)] = [f(x)][g(x)] = \varphi(f(x))\varphi(g(x))$$

Now consider the map $\phi : F \xrightarrow{i} F[x] \xrightarrow{\varphi} F[x]/(p(x))$ given by

$$a \mapsto a + 0x^1 + \ldots \mapsto [a]$$

The first map takes $a \in F$ to the constant polynomial $a \in F[x]$. The second map is the one just discussed. To see that ϕ is injective we compute $\ker \phi$. If $0 = \phi(a) = [a]$ then $p(x)|a(x)$. However as $a \in F$ if $a \neq 0$ then $\deg a = 0$ and $\deg p(x) \geqslant 1$. This is a contradiction. Thus ϕ is isomorphic onto its image $\phi(F)$. We set $F^* = \phi(F)$.

□

Why think of F as being identified with its image $\phi(F)$. In this way, we can think of F as a subring of $F[x]/(p(x))$. See Page 132 of BH for additional discussion on this.

Theorem 5.14. *Let* F *be a field and* $p(x)$ *a nonconstant polynomial in* $F[x]$*. If* $f(x) \in F[x]$ *and* $(f(x), p(x)) = 1$ *then* $[f(x)]$ *is a unit in* $F[x]/(p(x))$*.*

Proof. If $(f(x), p(x)) = 1$ then $1_F = u(x)f(x) + v(x)p(x)$ by Bezout's theorem. Therefore $1_F - u(x)f(x) = v(x)p(x)$. Passing to the congruence class we have $[1_F] = [u(x)f(x)] = [u(x)][f(x)]$. Thus $[f(x)]$ is a unit. □

The Structure of $F[x]/(p(x))$

Theorem 5.15. *Let* F *be a field and* $p(x)$ *a constant polynomial in* $F[x]$*. Then the following statements are equivalent:*

(1) $p(x)$ *is irreducible in* $F[x]$
(2) $F[x]/(p(x))$ *is a field*
(3) $F[x]/(p(x))$ *is an integral domain*

Proof.
(1) (1) $\implies$ (2). We've seen that $F(x)/(p(x))$ is a commutative ring with identity (Theorem 5.7 in BH). We must show it is a field. Consider $[f(x)] \neq 0$. Let's consider $f(x) \in F[x]$. Since p is irreducible $(p(x), f(x))$ is 1_F or an associate of $p(x)$. However, it cannot be an associate of $p(x)$ as $[f(x)] \neq 0$. Thus $(p(x), f(x) = 1_F$ and we invoke Theorem 5.14 to get $[f(x)]$ is a unit.
(2) (2) $\implies$ (3). We have already seen that every field is an integral domain in Theorem 3.8 of BH.
(3) (3) $\implies$ (1). Assume $p(x) = a(x)b(x)$ for nonzero $a(x), b(x)$. We want to show that $a(x), b(x)$ are units or associates. We have that $0 = [p(x)] = [a(x)][b(x)]$. Thus $[a(x)] = 0$ or $[b(x)] = 0$ since $F[x]/(p(x))$ is an integral domain. If $[a(x)] = 0$. Then $a(x) = a \in F$ and a is a unit. If $[b(x)] = 0$ then $b(x) = b \in F$ and a is an associate. Thus any divisor is either unit or associate.

□

Let F be a field and $p(x)$ irreducible in $F[x]$. Let $K = F[x]/(p(x))$. We've considered how F may be thought of as a subfield of $F[x]/(p(x))$. That is, F is a subfield of K. One also says that K is an *extension field* of F.

Since $F \subset K$, polynomials in $F[x]$ can be considered to have coefficients in the larger field K. We can ask about the roots of such polynomials in K.

Let's take an example. Consider $p(x) = x^2 + x + 1$ in $\mathbb{Z}_2[x]$. $p(x)$ has no roots in $\mathbb{Z}_2[x]$ (notice $f(1) = f(0) = 1$) therefore it is irreducible (Corollary 4.19 in BH). Therefore we can apply Theorem 5.15 to see that $K = \mathbb{Z}_2[x]/(x^2 + x + 1)$ is an extension field of $\mathbb{Z}_2$. Let's now test for roots in K, for instance [x]. Let's compute with [x], we see

$$[x]^2 + [x] + 1 = [x + 1] + [x] + 1 = [1] + [1] = [0]$$

We are considered the polynomial $x^2 + x + 1$ as residing in $K[x]$. We are then taking elements of K, such as [x], and inputting them into $p(x)$. This is a bit confusing and is perhaps cleared up by using a different notation. Say, $\alpha := [x] \in K$. Then $p(\alpha) = \alpha^2 + \alpha + 1 = 0$. Thus α is a root of $p(x)$ in $K[x]$. See discussion on Page 136 of BH.

Theorem 5.16. *Let* F *be a field and* $p(x) \in F[x]$ *irreducible. Then* $F[x]/(p(x))$ *is an extension field of* F *that contains a root of* $p(x)$.

Proof. Let $K = F[x]/(p(x))$. Then K is an extension field of F as we've seen. Let $p(x) = a_n x^n + \ldots + a_1 x + a_0$ with $a_i \in F$. Since $a_i \in F$ we have $a_i \in K$. Let $\alpha = [x]$ in K. We can compute (using our results about congruence class arithmetic)

$$\begin{aligned} a_n\alpha^n + \ldots + a_1\alpha + a_0 &= a_n[x]^n + \ldots + a_1[x] + a_0 \\ &= [a_n x^n + \ldots a_1 x + a_0] \\ &= [p(x)] = 0_F \end{aligned}$$

This last equality follows since $p(x) \equiv 0 \mod p(x)$. Therefore $\alpha \in K$ is a root of $p(x)$. □

Corollary 5.17. *Let* F *be a field and* $f(x) \in F[x]$ *be nonconstant. Then there is an extension field of* K *of* F *that contains a root of* $f(x)$.

Proof. Let $p(x)$ be an irreducible factor of $f(x)$. Thus we can consider $K = F[x]/(p(x))$. We have that K contains a root of $p(x)$. Every root of $p(x)$ is a root of $f(x)$, since $f(x) = p(x)g(x)$. Thus K contains a root of $f(x)$. □

Let's consider an example. We know that $p(x) = x^2 + 1$ is irreducible in $\mathbb{R}[x]$. Thus $p(x)$ has a root in $\mathbb{R}[x]/(x^2 + 1)$, namely $\alpha = [x]$. Thus $\alpha^2 = -1$. This sounds like the complex numbers $\mathbb{C}$.

See page 137 in BH for further discussion.

In this chapter we will generalize the idea of congruence to arbitrary rings. We have done congruence now in both $\mathbb{Z}$ and $F[x]$ - we will axiomatize the idea by writing down the common properties. This will result in the definition of an *ideal*.

We will observe that subring is not quite the correct notion for a 'subobject' of a ring, again leading to the development of the idea of *ideal*. We will prove very powerful structure theorems (these will also be called isomorphism theorems). This chapter contains very deep abstract algebra ideas. See this endnote for some meta-mathematics/mathematical philosophy.[1]

It is important to discriminate between *objects*, e.g. rings, and their *morphisms* (i.e a function between rings), e.g. ring homomorphism. One then talks of a *category* of rings (i.e. the collection of all rings) with morphisms - so the only functions you allow between rings are ring homomorphisms. More examples of objects include fields, groups[17], vector spaces. Their morphisms are, respectively, ring homomorphisms, group homomorphisms and linear maps. The last seventy years of **category theory**[2] have shown that *the morphisms are of much more importance than the objects themselves*. Proper 'subobjects', e.g. a subring, should arise somehow through some appropriate morphism. In particular, as the kernel of some morphism. In this chapter we'll see that our definition of subring does not fit this description.[18] We'll introduce the notion of ideal instead, and show that these do arise as kernels.

6. Ideals and Congruence (and Morphisms)

> *My [algebraic] methods are really methods of working and thinking; this is why they have crept in everywhere anonymously.*
>
> – Emmy Noether[19]

Recall that in the ring $\mathbb{Z}$ the notation $a \equiv b \mod 3$ means that $a - b$ is divisible 3. Let $I = \{0, \pm 3, \pm 6, \pm 9, \dots\}$.

Then congruence can be characterized as follows:

$$a \equiv b \mod 3 \qquad \text{means} \qquad a - b \in I$$

Observe that I is a subring of $\mathbb{Z}$ (closed under subtraction, multiplication).

Furthermore, the product of any integer $n \in \mathbb{Z}$ is itself a multiple of 3. Thus the subring I has the property

$$\text{for } k \in \mathbb{Z} \text{ and } i \in I \text{ we have } ki \in I$$

This was of course similar to congruence in the polynomial ring. The notation $f(x) \equiv g(x) \mod (x^2 - 2)$ in the polynomial ring implies that $f(x) - g(x)$ is a multiple of $x^2 - 2$. Let

$$I = \{h(x)(x^2 - 2) : h(x) \in \mathbb{Q}[x]\}$$

I is a subring of $\mathbb{Q}[x]$ which has the following property:

$$\text{for } k(x) \in \mathbb{Q}[x] \text{ and } t(x) \in I \text{ we have } k(x)t(x) \in I$$

[17] We'll come to these in Chapter 7

[18] Can you find a subring that is not a kernel of some homomorphism? On its face may seem like a hard problem. We'll may able to do this by the end of this Chapter.

[19] Letter to H. Hasse (1931). This can be found in *Kleiner, I. (2007). A history of abstract algebra. Springer Science & Business Media.*

Therefore congruence modulo $x^2 - 2$ may be described in terms of I:

$$f(x) \equiv g(x) \mod x^2 - 2 \qquad \text{means} \qquad f(x) - g(x) \in I$$

This suggests that congruence in a ring may be defined in terms of certain subrings.

Definition 6.1. Let $I \subset R$ be a subring. I is called an *ideal* If for all $r \in R$ and $a \in I$ then $ra \in I$ and $ar \in I$.

Remark 6.2. *This property is sometimes called the* **absorption property** *as the product of any arbitrary element and ideal element lands by in the ideal (i.e. is absorbed).*

Here are two easy examples of ideals: the zero ideal $\{0_R\}$ and the entire ring R.

Consider the ring $\mathbb{Z}[x]$. Let I be the set of polynomials whose constant terms are even integers, i.e. $I = \{a_n x^n + \ldots + 2a_0 \in \mathbb{Z}[x] : a_i \in \mathbb{Z}\}$. Thus $x^3 + x + 6$ is in I, but $4x^2 + 3$ is not. Why is I an ideal?

We can that I is a subring: How? (hint: is it a kernel?) For any $k(x) \in \mathbb{Z}[x]$ and $t(x) \in \mathbb{Z}[x]$ we have $k(x)t(x) \in I$ since the constant term on this product must be even (this follows from polynomial multiplication).

Let's do more examples. Let T be the ring of all functions from $\mathbb{R}$ to $\mathbb{R}$. Let $I = \{f \in T : f(2) = 0\}$. Then I is an ideal. Why?

Here is a *nonexample*. Consider the subring $\mathbb{Z} \subset Q$. Notice $\mathbb{Z} \subset \mathbb{Q}$ is not an ideal. Why? Consider $\frac{1}{2}5$. Therefore there are subrings which are *not* ideals.

Theorem 6.3. *A nonempty subset* I *of a ring* R *is an ideal if and only if it has these properties:*

(1) *if* $a, b \in I$ *then* $a - b \in I$
(2) *if* $r \in R$ *and* $a \in I$ *then* $ra \in I$ *and* $ar \in I$

Proof. Every ideal must have these properties. Now assume that $I \subset R$ has these properties. We wish to show that I is an ideal. Then I absorbs products by (2). Furthermore, (2) implies I is closed under multiplication. We also have that by (1) that I is closed under subtraction. □

Theorem 6.4. *Let* R *be a commutative ring with identity,* $c \in R$. *Let* I *be the set of all multiplies of* c *in* R, *that is,* $I = \{rc : r \in R\}$. *Then* I *is an ideal.*

Proof. If $r_1, r_2 \in R$ and $r_1c, r_2c \in I$, then

$$r_1c - r_2c = (r_1 - r_2)c \in I$$

and

$$r(r_1c) = (rr_1)c \in I \qquad \text{and} \qquad (r_1c)r = rr_1c \in I$$

□

An ideal of the form $I = \{rc : r \in R\}$ is called a *principal ideal generated by* c and is denoted (c). For example, $(3) \subset \mathbb{Z}$ is the ideal generated by 3 (all multiples of 3).

In any commutative ring with identity, the principal ideal (1_R) is the entire ring since $r = r1_R \in (1_R)$. The *principle* refers to the fact that such an ideal is generated by only one element. Ideals need not be principal.

Here's an example. Let I be the set of polynomials in $\mathbb{Z}[x]$ with even constant terms.[20] We claim that I is not principal. To see this, assume that I was principal, ie. $I = (p(x))$. Then since $2 \in I$ we have $2 = r(x)p(x)$. Thus $0 = \deg 2 = \deg r(x) + \deg p(x)$. Therefore $\deg p(x) = 0$. This implies $p(x) = c \in 2\mathbb{Z}$. Since $c|2$ we have $c = \pm 1$. Since $x + 0 \in I$ we have $x = g(x)p(x)$. This forces $\deg g(x) = 1$, thus $g(x) = ax + b$. So $\pm 2(ax + b) = x$. This forces $\pm 2a = 1$, with $a \in \mathbb{Z}$. This is impossible. Thus I is not principal.

We generalize principal (generated by one element) with the following definition:

Theorem 6.5. *Let* R *be a commutative ring with identity and* $c_1, c_2, \ldots c_n \in R$*. The set* $I = \{r_1c_1 + r_2c_2 + \ldots + r_nc_n : r_i \in R\}$ *is an ideal in* R*.*

Proof. We first show I is a subring. I is nonempty since $c_1 \in I$. Let $a, b \in I$. Then $a = r_1c_1 + \ldots r_nc_n, b = s_1c_1 + \ldots + s_nc_n$ and $a - b = (r_1 - s_1)c_1 + \ldots (r_n - s_n)c_n \in I$. We have $ab = (r_1c_1 + \ldots + r_nc_n)(s_1c_1 + \ldots + s_nc_n)$. Using commutativity and distributivity, it is straightforward that this is in I. Finally let $r \in R$. Then $ra = rr_1c_1 + \ldots + rr_nc_n \in I$. □

We've see congruence in $\mathbb{Z}$ and $F[x]$. Now we define it for an arbitrary ring.

Definition 6.6. *Let* $I \subset R$ *be an ideal. Let* $a, b \in R$*. Then* a is congruent to b modulo I, $a \equiv b \mod I$, *if* $a - b \in I$.

Let T be the ring of functions $\mathbb{R} \to \mathbb{R}$. Let $I = \{g \in T : g(2) = 0\}$. Consider $f(x) = x^2 + 6$ and $h(x) = 5x$. Then $(f - h)(2) = f(2) - h(2) = 2^2 + 6 - 10 = 0$. Thus $f - h \in I$, so $f \equiv h \mod I$.

We'll now see that congruence modulo I is an equivalence relation, as we've done for $\mathbb{Z}$ and $F[x]$.

Theorem 6.7. *Let* I *be an ideal in* R*. Then the relation of congruence modulo* I *is*

(1) *Reflexive.* $a \equiv a \mod I$ *for every* $a \in R$
(2) *Symmetric.* $a \equiv b \mod I$ *then* $b \equiv a \mod I$.
(3) *Transitive.* $a \equiv b \mod I$ *and* $b \equiv c \mod I$ *implies* $a \equiv c \mod I$.

Proof. (1) $a - a = 0_R \in I$
(2) If $a - b \in I$ then $-(b - a) \in I$ since I is a subring.
(3) We have $a - b \in I$ and $b - c \in I$. Thus $a - c = (a - b) + (b - c) \in I$ since I is closed under addition.

□

Here's another that we have seen before:

Theorem 6.8. *Let* $I \subset R$ *be an ideal. If* $a \equiv b \mod I$ *and* $c \equiv d \mathrm{mod} I$ *then*

(1) $a + c \equiv b + d \mod I$
(2) $ac \equiv bd \mod I$

Proof. We have $a - b, c - d \in I$.

[20] An elegant (if perhaps complicated) way to express I is as the kernel of the composition of homomorphisms

$$\mathbb{Z}[x] \xrightarrow{\varphi} \mathbb{Z} \xrightarrow{\psi} \mathbb{Z}_2$$

Here φ is the evaluation map $p(x) \mapsto p(0)$. Thus maps $p(x)$ to its constant term. The map $\mathbb{Z} \to \mathbb{Z}_2$ sends $n \mapsto [n]_2$. Thus $\ker(\psi \circ \varphi)$ is the set of polynomials with even constant term. We saw this on Exam 1. It is a good idea to get comfortable with compositions, kernels and images.

(1) $(a+c)-(b+d)=(a-b)+(c-d)\in I$
(2) $ac-bd=ac-bc+bc-bd=(a-b)c+b(c-d)$. $(a-b)c$ and $b(c-d)$ are both in I (absorption), thus their sum is in I.

□

If I is an ideal in R and $a\in R$ then *congruence class of* a *modulo* I is the set of all elements of R that are congruent to a modulo I. We compute it in the fashion we have before:

$$\begin{aligned}\{b\in R: b\equiv a \mod I\}&=\{b\in R: b-a\in I\}=\{b\in R: b-a=i, i\in I\}\\&=\{b\in R: b=a+i, i\in I\}\\&=\{a+i: i\in I\}\end{aligned}$$

This time, we'll denote the congruence class of a modulo I by the symbol $a+I$ rather than $[a]$. This change of notation is justified by the fact that $a+I=\{a+i: i\in I\}$.[21] $a+I$ is called a *coset* of I in R.

Theorem 6.9. *Let* I *be an ideal in* R *and let* $a,c\in R$. *Then* $a\equiv c \mod I$ *if and only if* $a+I=c+I$.

Proof. Let $x\in a+I$. Then $x=a+i$ for some $i\in I$. We have $a-c=i'$ for $i\in I$. Thus $x=(c+i')+i=c+(i'+i)\in c+I$. A similar argument shows $c+I\subset a+I$

Assume $a+I=c+I$. Then $a\in a+I=c+I$. Thus $a=c+i$, so $a-c=i\in I$. Thus $a\equiv c \mod I$. □

Another proof we've seen before:

Corollary 6.10. *Let* $I\subset R$ *be an ideal. Then two cosets of* I *are either disjoint or equal.*

Let's consider $I=(3)$. The cosets of I are the congruence classes modulo 3. Thus $0+I, 1+I, 2+I$. So the set $\mathbb{Z}/I$ of cosets is the set $\mathbb{Z}_3$.

Consider again $I\subset\mathbb{Z}[x]$ the ideal of all polynomials with even constant term. It is straightforward that $\mathbb{Z}[x]/I$ consists of two distinct cosets: for $f(x)\in\mathbb{Z}[x]$ we have $f(x)$ has constant term either even or odd. If even then $f(x)\in I$ and if odd then $f(x)\in 1+I$.

Consider $T=\{f:\mathbb{R}\to\mathbb{R}\}$. Let $I=\{f\in T: f(2)=0\}$. Then $f-g\in I$ if and only if $f(2)-g(2)=0$. Consider the constant function $h(x)\equiv r\in R$. $h(x)$ is in its unique coset.

Let's relate ideals and homomorphisms. Here are two lemmas we've seen before for subrings.

Lemma 6.11. *Let* $f:R\to S$ *be a homomorphism. Let* $T\subset S$ *be an ideal. Then* $f^{-1}(T)$ *is an ideal in* R.

Proof. We have from previous results that $f^{-1}(T)$ is a subring. We just have to show that absorption property.

Let $x\in R$ and $a\in f^{-1}(T)$. We want to show $xa\in f^{-1}(T)$. To check this, apply f. We have $f(xa)=f(x)f(a)\in f^{-1}(T)$ since $f(a)\in T$ and T is an ideal. Thus $xa\in f^{-1}(T)$. Similarly, $ax\in f^{-1}(T)$. Thus $f^{-1}(T)$ is an ideal. □

Lemma 6.12. *Let* $f:R\to S$ *be a surjective homomorphism. Let* $I\subset R$ *be an ideal. Then* $f(I)$ *is an ideal in* S.

[21] Notice that $a+I$ is just a formal notation. It does not indicate any real 'addition'. We haven't defined the sum of an element $a\in R$ and an ideal $I\subset R$.

Proof. We've seen that $f(I) = im(f) = \{f(a)\ a \in R\}$ is a subring. We now show absorption. Let $x \in S$ and $y \in f(I)$. Since f is surjective, there are $a \in R$ and $b \in I$ with $f(a) = x, f(b) = y$. Thus $xy = f(a)f(b) = f(ab)$. $b \in I$ thus $ab \in I$. Therefore $xy = f(ab) \in f(I)$. □

Remark 6.13. *You should understand precisely why/when surjectivity is needed for Lemma 6.12!*

Remark 6.14. *Compare and contrast these results to what we know about subrings!*

Quotient Rings and Homomorphisms

We proceed as we've done before, and show that the congruence classes modulo an ideal form a ring.

Let I be an ideal in a ring R. The elements of R/I are the cosets of I, i.e. $a + I := \{a + i : i \in I\}$. We'll show that R/I forms a ring by defining a multiplication and addition operation on the cosets themselves. To do this we use arithmetic mod I.

Theorem 6.15. *Let $I \subset R$ be an ideal. If $a + I = b + I$ and $c + I = d + I$ in R/I, then*

$$(a + c) + I = (b + d) + I \qquad \textit{and} \qquad ac + I = bd + I$$

Proof. Straightforward from what we've done before. □

We define addition and multiplication as we've done before on congruence classes.

$$(a + I) + (c + I) = (a + c) + I$$

and

$$(a + I)(c + I) = ac + I$$

Here '$+$' is being used for three different meanings. First, as the formal symbol denoted $a + I$. Second, as an operation of elements of R with $a + c$. Third, as an addition operation on cosets, that is, the operation being defined.

Here's an example. Let F be a field, $p(x) \in F[x]$ and $I = (p(x))$. The cosets of I are the congruence classes modulo $p(x)$. Therefore $F[x]/I$ is the ring $F(x)/(p(x))$.

Theorem 6.16. *Let $I \subset R$ be an ideal. Then*

1. R/I *is a ring (with addition, multiplication as defined previously)*
2. *If* R *is commutative then* R/I *is commutative*
3. *If* R *has an identity, then so does* R/I

Proof. (1) We'll leave this for an exericse.
(2) If R is commutative then $(a + I)(c + I) = ac + I = ca + I = (c + I)(a + I)$
(3) What is the identity in R/I? (How about $1_R + I$?)

□

The ring R/I is called the *quotient ring* of R by I. One way to think about R/I is that you are setting the elements in I to zero, i.e. you are collapsing the ideal I to zero. This is what it means to work 'mod I'.

6.1. **Homomorphisms.** Quotient rings are the generalization of congruence class arithmetic in $\mathbb{Z}$ and $F[x]$. We dabbled a bit with the relationship between homomorphisms and quotients before, but in this section we will formalize this.

Recall that definition of *kernel* of a ring homomorphism $f : R \to S$

$$\ker f := \{r \in R : f(r) = 0_S\}$$

An equivalent way to express the kernel is as the fiber over 0_S, i.e. $\ker f = f^{-1}(0_S)$.

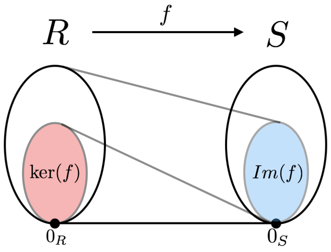


FIGURE 1. This is a good image to have in mind for how kernels and images work.

Consider the homomorphism $f : \mathbb{Z} \to \mathbb{Z}_6$ defined by $f(r) = [r]_6$. An easy computation gets the kernel

$$\ker f = \{r \in \mathbb{Z} : f(r) = [0]\} = \{r \in \mathbb{Z} : [r] = [0]\} = \{r \in \mathbb{Z} : r \equiv 0 \mod 6\} = \{6k : k \in \mathbb{Z}\} = (6)$$

Consider the evaluation homomorphism $\varphi : \mathbb{R}[x] \to \mathbb{R}$ sending $p(x) \mapsto p(0)$, i.e. the constant term. The kernel consists of all polynomials with constant term 0. All such polynomials are divisible by x, thus $\ker \varphi = (x)$.

Theorem 6.17. *Let $f : R \to S$ be a homomorphism of rings. Then $\ker f$ is an ideal in R.*

Proof. We've seen before that the kernel is a subring. Now we must show the absorption property. Let $r \in R$ and $a \in \ker f$. Then

$$f(ra) = f(r)f(a) = 0_S = f(a)f(r) = f(ar)$$

Thus $ra, ar \in \ker f$.

□

Remark 6.18. *We've thus shown a relationship between kernels and ideals. In particular, any kernel is an ideal. Moreover, we will see that an ideal I is the kernel of the epimorphism $R \to R/I$ given by $r \mapsto r + I$.*

Recall the following theorem, which we have seen before.

Theorem 6.19. *Let $f : R \to S$ be a homomorphism of rings. Then $\ker f = \{0_R\}$ if and only if f is injective.*

Proof. See Page 156 in BH. Or earlier in these notes. □

The following theorem is *of fundamental importance* to understanding the relationship between kernels and quotient rings. We've studied the rings R and and quotient ring R/I. For instance $\mathbb{Z}$ and $\mathbb{Z}_2$. One can consider these as two separate rings. However it is much more important to understand that these are related by a particular homomorphism,

$n \mapsto [n]_2$. This homomorphism captures the structure of how these two rings are related. In particular it shows how $\mathbb{Z}_2$ can be 'obtained' from $\mathbb{Z}$.

Theorem 6.20 (Natural Homomorphism)**.** *Let* I *be an ideal in* R*. The map* $\pi : R \to R/I$ *given by* $\pi(r) = r + I$ *is a surjective homomorphism with kernel* I

Proof. The map π is surjective since a coset $r + I$ is in the image of $\pi(r)$. The definition of addition and multiplication in R/I shows that π is a homomorphism

$$\pi(r + s) = (r + s) + I = (r + I) + (s + I) = \pi(r) + \pi(s)$$

and

$$\pi(rs) = rs + I = (r + I)(s + I) = \pi(r)\pi(s)$$

$\pi(r) = r + I = I$ if and only if $r \in I$. Thus $\ker f = I$. □

Remark 6.21. *The notation* π *is conventional notation for a quotient map such as this. The map* π *is called the* **natural homomorphism***.*

The homomorphism $\pi : R \to R/I$ is a special case of a more general situation; the situation of a surjective homomorphism $f : R \to S$. In this case S is called the *homomorphic image* of R, since $\mathrm{im}(f) = f(R) = S$. If f is an isomorphism, we know that R and S have the same structure. In particular, if f is an isomorphism, it is injective, and thus $\ker f = \{0_R\}$. If f is not an isomorphism, properties in one ring may not hold in the other. Even if f is not an isomorphism, we the properties of S and fact that f is a homomorphism give us information about R. BH provides a nice visual analogy:

> *[Think of a sculpture/shape* R *in* $\mathbb{R}^3$*.] If* $f : R \to S$ *is an isomorphism then* S *is an exact, three dimensional replica of* R*. If* f *is only a surjective homomorphism, then* S *is a two-dimensional photographic image of* R*, in which some features of* R *are accurately reflected but others are distorted or missing.*

Let's do some visual examples. The first is Figure 2, which describes a relationship between the ring $\mathbb{R}$ and the ring $\mathbb{R}/I$ where $I = (2\pi)$. $\mathbb{R}/(2\pi)$ is often thought of as the circle.[22]

To quote again from BH, because this is a very nice description:

> *The following theorem states that* every *homomorphic image of a ring* R *is isomorphic to a quotient ring* R/K *for some ideal* K*. Thus, if you know all of the quotient rings of* R*, then you know all the possible homomorphic images of* R*. The ideal* $K = \ker f$ *measures how much information is lost in passing from the ring* R *to the homomorphic image* R/K*. When* $\ker f = (0_R)$ *then* f *is an isomorphism.*

This is an extremely important structure theorem.

Theorem 6.22 (First Isomorphism Theorem)**.** *Let* $f : R \to S$ *be a surjective homomorphism. Let* $K = \ker f$*. Then the quotient ring* R/K *is isomorphic to* S*.*

Proof. We define a function from $\varphi : R/K \to S$ and show its an isomorphism. Define $\varphi(a + K) = f(a)$. First of all we must show this is well defined, i.e. for $a + K = b + K$ we have $\varphi(a + K) = \varphi(b + K)$. If $a + K = b + K$ then $a - b \in K$, thus $a - b \in \ker f$. Therefore

[22] This construction is natural in *topology*. In topology it is called a *covering* of the circle. See https://en.wikipedia.org/wiki/Covering_space for more.

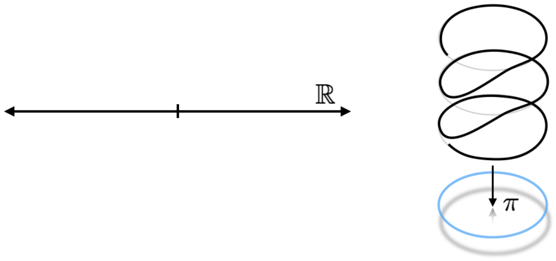


FIGURE 2. Let's consider the epimorphism $p : \mathbb{R} \to \mathbb{R}/(2\pi)^{\dagger}$ given by

$$\mathbb{R} \ni x \mapsto [x]_{2\pi}$$

We think of $\mathbb{R}$ as the real axis. We can think of $R/(2\pi)$ as the circle (Why?). We can visualize the structure as $\mathbb{R}$ wrapping around a 'cylinder' along the z-axis, then the map p is a projection down onto the plane. What is $\ker p$? What is the fiber of p above $[\pi]$?

† Unfortunately in the figure the projection is drawn as $\pi : R \to R/(2\pi)$. In the text we use p vs. π for the natural homomorphism as we want to discuss $\pi \in \mathbb{R}$ in this example.

$0 = f(a-b) = f(a) - f(b) = \varphi(a+K) - \varphi(b+K)$. If $s \in S$ then there exists some $a \in R$ with $f(a) = s$. Thus $\varphi(a+K) = f(a) = s$. So φ is surjective. To see that φ is injective, notice that $\varphi(a+K) = f(a) = 0_S$ implies that $a \in \ker f$. Thus $a \in K$. Finally we show that φ is a homomorphism. We compute

$$\varphi[(a+K)+(b+K)] = \varphi[(a+b)+K] = f(a+b) = f(a)+f(b) = \varphi(a+K)+\varphi(b+K)$$

and

$$\varphi[(a+K)(b+K)] = \varphi[ab+K] = f(ab) = f(a)f(b) = \varphi(a+K)\varphi(b+K)$$

Thus φ is an isomorphism.

□

Remark 6.23. *A slightly more general way to state this theorem is that* $R/\ker f \cong \operatorname{Im}(f)$ *for a ring homomorphism* $f : R \to S$.

Remark 6.24. *This result is sometimes known as 'Noether's Isomorphism Theorem' and is generally credited to Emmy Noether.*

Look to Figure 3 for a picture of the theorem.

Let's do some examples. In the ring $\mathbb{Z}[x]$ the principal ideal $(x) = \{k(x)x : k(x) \in \mathbb{Z}[x]\}$, all polynomials with constant term 0. A natural question: what does the quotient ring $\mathbb{Z}[x]/(x)$ look like? We can answer this question but using the evaluation homomorphism. Consider the map $\varphi : \mathbb{Z}[x] \to \mathbb{Z}$ given by $p(x) \mapsto p(0)$, i.e. map the polynomial to its constant term. What is $\ker \varphi$? It is precisely the set of polynomials with constant term 0, which is (x). Since φ is surjective (as we've discussed before) and $\ker \varphi = (x)$ we have $\mathbb{Z}[x]/(x) = \mathbb{Z}[x]/\ker \varphi \cong \mathbb{Z}$.

Consider $T = \{f : \mathbb{R} \to \mathbb{R}\}$ and $I = \{f \in T : f(2) = 0\}$. What is T/I? If we can cook up an appropriate homomorphism, then we can use the First Isomorphism Theorem.

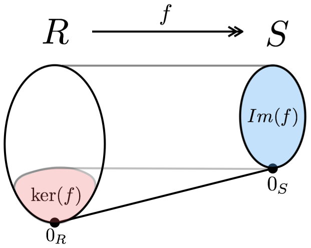


FIGURE 3. Visual Interpretation of the First Isomorphism Theorem. The '↠' is notation for a surjective homomorphism (epimorphism). Can you see why this image is a (albeit crude) depiction the fact $R/\ker f \cong S$? Notice here that we are collapsing $\ker f$ to zero.

Let's define $\varphi : T \to \mathbb{R}$ by $\varphi(f) = f(2)$.[23] Then φ is surjective and a homomorphism as $\varphi(f+g) = (f+g)(2) = f(2) + g(2) = \varphi(f) + \varphi(g)$ and $\varphi(fg) = (fg)(2) = f(2)g(2) = \varphi(f)\varphi(g)$. Then $\ker \varphi = \{g \in T : g(2) = 0\} = I$. Thus $T/I = T/\ker\varphi \cong \mathbb{R}$.

Finally, what do homomorphic images of $\mathbb{Z}$ look like? Suppose $f : \mathbb{Z} \to S$ is a homomorphism. If f is an isomorphism then S is an 'isomorphic copy' of $\mathbb{Z}$. If f is surjective, then $\mathbb{Z}/\ker f \cong S$. Since $\ker f$ is an ideal in $\mathbb{Z}$, it is principal (this is a fairly straightforward argument[24]), so $\ker f = (n)$ for some $n \neq 0$. Thus $S \cong \mathbb{Z}/\ker f = \mathbb{Z}/(n) = \mathbb{Z}_n$. Thus every homomorphic image of $\mathbb{Z}$ is either isomorphic to $\mathbb{Z}$ or $\mathbb{Z}_n$ for some n.

THE STRUCTURE OF R/I WHEN I IS PRIME OR MAXIMAL

Primes in $\mathbb{Z}$ and irreducibles in $F[x]$ play the same role in the structure of the respective quotient rings. We now want to abstract this concept to general commutative rings.

Recall that in $\mathbb{Z}$ we have that p is prime if and only if $p \in \mathbb{Z}$ such that if $p|bc$ then $p|b$ or $p|c$. To say $p|b$ means that b is a multiple of p. In terms of ideals this means that $b \in (p)$.

Thus we can say that p is prime if and only if whenever $bc \in (p)$ then $b \in (p)$ or $c \in (p)$. This motivates our following, abstract definition.

Definition 6.25. An ideal P in a commutative ring R is said to be *prime* if $P \neq R$ and if $bc \in P$ then $b \in P$ or $c \in P$.

Let's go back to $\mathbb{Z}$ for an example. The principal ideal (p) is prime in $\mathbb{Z}$ whenever p is prime. On the other hand (6) (the ideal) is not prime, since $2 \cdot 3 = (6)$ but $2, 3 \notin (6)$.

Here's a subtle difference between prime integers and prime ideals: (0) in $\mathbb{Z}$ is prime since $ab = 0$ implies $a = 0$ or $b = 0$.

Let's do examples in $\mathbb{Z}[x]$:

Let $I = \{a_n x^n + \ldots + a_1 x + 2a_0 : a_i \in \mathbb{Z}\} \subseteq \mathbb{Z}[x]$. Then $I \neq \mathbb{Z}[x]$. Let $f(x) = a_n x^n + \ldots + a_0$ and $g(x) = b_m x^m + \ldots b_0$ such that $f(x)g(x) \in I$. The constant term of $f(x)g(x)$ is $a_0 b_0$ and is even. Thus a_0 or b_0 must be even since the product of two odd integers is odd. This implies $f(x)$ or $g(x)$ is in I. Thus I is prime.

[23] That is $f \mapsto f(2)$. This is an evaluation homomorphism for general functions instead of just polynomials

[24] Can you show this? Use the Well Ordering Principle and Bezout's Theorem

Recall that $(x) \subseteq \mathbb{Z}[x]$ is the principal ideal generated by x, i.e. the polynomials with zero constant term. Is (x) prime? If $f(x) = a_n x^n + \dots a_0$ and $g(x) = b_m x^m + \dots + b_0$ and $f(x)g(x) \in (x)$ then $a_0 b_0 = 0$. In this case $a_0 = 0$ or $b_0 = 0$, thus $f(x) \in (x)$ or $g(x) \in (x)$.

We've computed $\mathbb{Z}[x]/(x) \cong \mathbb{Z}$. This shows that R/P is not necessarily a field if P is prime (this contrasts $\mathbb{Z}/(p) = \mathbb{Z}_p$ being a field when p is prime). The best we can get with prime ideals is that R/P is an integral domain, as the next result shows.

Remark 6.26. *We'll do this proof a bit differently than done in BH. We want to stress the importance of the natural homomorphism $\pi : R \to R/P$. See Theorem 6.14 in BH for alternate proof.*

Theorem 6.27. *Let P be an ideal in a commutative ring R with identity. Then P is a prime ideal if and only if the quotient ring R/P is an integral domain.*

Proof. Let $\pi : R \to R/P$ be the natural morphism, i.e. $a \mapsto a + P$. Note that $\ker \pi = P$.

Let P be prime. Any $a + P$ in R/P can be written as $\pi(a)$. Let $\pi(a)\pi(b) = 0$. We show $\pi(a) = 0$ or $\pi(b) = 0$. We have $0 = \pi(a)\pi(b) = \pi(ab)$. Thus $ab \in \ker \pi = P$. Since P is prime $a \in P$ or $b \in P$. Since $P = \ker \pi$ this implies either $\pi(a) = 0$ or $\pi(b) = 0$. It remains to show $1_R + P \neq 0_R + p$. Since P is prime $P \neq R$, thus $1_R \notin P$ so $1_R + P \neq P$.

Now let R/P be an integral domain. Assume $ab \in P$. We show $a \in P$ or $b \in P$. We have $0 = \pi(ab) = \pi(a)\pi(b)$. Thus $\pi(a) = 0$ or $\pi(b) = 0$. This implies $a \in P$ or $b \in P$. It remains to show $P \neq R$. Since R/P is an integral domain $1_R + P \neq P$ thus $1_R \notin P$. So $P \neq R$.

□

Since a R/P for P prime is not necessarily a field, it is natural to ask what condition on P guarantees that R/P is a field. We now answer that question.

Definition 6.28. An ideal M is *maximal* if $M \neq R$ and whenever J is an ideal such that $M \subset J \subset R$ then $M = J$ or $J = R$.

Let's consider some examples. Is $(3) \subset \mathbb{Z}$ maximal? Notice that $\mathbb{Z}/(3) \cong \mathbb{Z}_3$ is a field.

Is $(x) \subset \mathbb{Z}[x]$ maximal? We saw that $\mathbb{Z}[x]/(x)$ is not a field.

Here are some nice relationships between fields and ideals.

Lemma 6.29. *Let F be a commutative ring with identity. Then F is a field if and only if the only ideals of F are 0 or F.*

Proof. Let F be a field. Let $I \subset F$ be an ideal. Then $1_F \in I$ or $1_F \notin I$. If $1_F \in I$ then $I = F$ (Why? for any $a \in F$ we have $a = a1_F \in I$). Suppose $1_F \notin I$. Let $a \in I$. If $a \neq 0$ then by absorption for $a^{-1} \in F$ we have $a^{-1}a = 1_F \in I$. This is a contradiction.

Now assume the only ideals of F are 0 or F. We show F is a field. Let $a \in F$ with $a \neq 0$. Consider $(a) = \{xa : x \in F\}$. $(a) \neq 0$ (why?) thus $(a) = F$. Thus there is x such that $xa = 1_F$. So F is a field. □

Theorem 6.30. *Let M be an ideal in a commutative ring R with identity. Then M is maximal if and only if the quotient ring R/M is a field.*

Proof. Consider the natural homomorphism $\pi : R \to R/M$ given by $a \mapsto a + M$. We know by Theorem 6.20 that π is a surjective homomorphism with $\ker \pi = M$.

Assume that M is maximal. We want to show that R/M is a field. We'll use Lemma 6.29. Suppose $J \subset R/M$ is an ideal. Then $\pi^{-1}(J)$ is an ideal by Lemma 6.11 and

$$M \subseteq \pi^{-1}(J) \subseteq R$$

Thus $\pi^{-1}(J) = M$ or $\pi^{-1}(J) = R$ by maximality of M. If $\pi^{-1}(J) = M$ then $J = 0$. If $\pi^{-1}(J) = R$ then $J = R/M$. Thus R/M is a field by Lemma 6.29.

Now assume that R/M is a field. We want to show that M is maximal. Let $M \subseteq J \subseteq R$. Then $0 = \pi(M) \subseteq \pi(J) = \pi(R) = R/M$ (Why?). $\pi(J)$ is an ideal of $p(R) = R/M$ by Lemma 6.12).[25] By Lemma 6.29 we have that $\pi(J) = 0$ or $\pi(J) = R$. If $\pi(J) = 0$ then $J = M$. If $\pi(J) = R/M$ then $J = R$. Thus M is maximal.

□

Remark 6.31. *See the proof in BH, Theorem 6.15 (pg 165) for a much more hands on approach! The approach presented here is (perhaps) more elegant. It illustrates the power of carrying around the natural homomorphism $\pi : R \to R/M$ instead of thinking of R and R/M as completely distinct objects. This proof hints at the depth of insight that can be gained by understanding the morphisms between two structures.*

Corollary 6.32. *In a commutative ring R with identity, every maximal ideal is prime.*

Proof. If M is a maximal ideal then R/M is a field. Thus R/M is an integral domain. Thus M is prime. □

The ideal I of polynomials with even constant terms in $\mathbb{Z}[x]$ is maximal since $\mathbb{Z}[x]/I$ is a field - it is isomorphic to $\mathbb{Z}_2$ as we have seen before.

Another example, let $T = \{f : \mathbb{R} \to \mathbb{R}\}$ and $I = \{f \in T : f(2) = 0\}$. We've seen that $T/I \cong \mathbb{R}$. Thus I is maximal.

We say in the proof of Theorem 6.30 that ideals in R/M and R are related. We now formalize this idea.

Proposition 6.33 (Correspondence Theorem)**.** *Let $I \subset R$ be an ideal. Consider the natural homomorphism $\pi : R \to R/I$. Set $\bar{R} = R/I$.*

1. *There is a bijective correspondence between the set of ideals of R which contain I and the set of all ideals of R/I. This correspondence is given by*
$$J \mapsto \pi(J) \qquad \pi^{-1}(\bar{J}) \mapsto J$$
2. *If $J \subset R$ corresponds to $\bar{J} \subset \bar{R}$ then R/J and $\bar{R}/\bar{J}$ are isomorphic rings.*

Proof. For ideal $\bar{J} \subset \bar{R}$ have by Lemma 6.11 that $\pi^{-1}(\bar{J})$ is an ideal. Since π is surjective, for ideal J we have $\pi(J)$ is an ideal of $\bar{R}$ by Lemma 6.12. This shows that the correspondence does send ideals to ideals. We must show that it is bijective. It suffices to show that $\pi(\pi^{-1}(\bar{J})) = \bar{J}$ and $\pi^{-1}(\pi(J)) = J$ (Why?). The equality $\pi(\pi^{-1}(\bar{J})) = \bar{J}$ holds since π is surjective. Furthermore $I \subset \pi^{-1}(\pi(J))$ holds for any map of sets. Let $x \in \pi^{-1}(\pi(J))$. We show $x \in J$. We have $\pi(x) \in \pi(J)$. Thus there exists $y \in J$ such that $\pi(y) = \pi(x)$. Now we have $0 = \pi(x) - \pi(y) = \pi(x - y)$. Thus $x - y \in \ker \pi = I$. Since $I \subset J$ and $y \in J$ we have $x = (x - y) + y \in J$.

[25] Be careful here. Note that homomorphic images of ideals are not necessarily ideals of the codomain. For ideal $I \subset R$ and homomorphism $f : R \to S$ the homomorphic image $f(I)$ is an ideal of the subring $f(R)$, not necessarily S.

Now let $I \subset J$. Denote $\bar{J} = \pi(J)$. Consider the composition

$$R \xrightarrow{\pi} \bar{R} \xrightarrow{\varphi} \bar{R}/\bar{J}$$

Here π is the natural homomorphism $R \to R/I$ and φ is the natural homomorphism $\bar{R} \to \bar{R}/\bar{J}$. Note that $\ker \varphi = \bar{J}$. Let $\phi := \phi \circ \pi$. Then $\phi : R \to \bar{R}/\bar{J}$. Since φ and π are surjective so is ϕ. Therefore by the First Isomorphism theorem we have that $R/\ker(\phi) \cong \bar{R}/\bar{J}$. Let's compute $\ker(\phi)$.

$$\ker(\phi) = \{a \in R : \phi(a) = 0\} = \{a \in R : \pi(a) \in \ker \varphi\} = \{a \in R : \pi(a) \in \bar{J}\} = \pi^{-1}(\bar{J})$$

Thus $R/\pi^{-1}(\bar{J}) \cong \bar{R}/\bar{J}$. From above we have $\pi^{-1}(\bar{J}) = \pi^{-1}(\pi(J)) = J$. Thus

$$R/J \cong \bar{R}\bar{J}$$

□

Remark 6.34. *Number* (2) *here is sometimes called the 'Third Isomorphism Theorem'.*

> *The introduction of the cipher 0 or the group concept was general nonsense too, and mathematics was more or less stagnating for thousands of years because nobody was around to take such childish steps.*
>
> – Alexander Grothendieck[26]

7. Group Theory

We've studied the rings $\mathbb{Z}, \mathbb{Z}_n, \mathbb{R}, F[x]$, etc. These have been sets equipped with *two* operations. Groups will be sets with only *one* operation. Some groups arise from rings by ignoring one of the operations. Many of the ideas, such as homomorphisms, kernels, images, cosets, will also be found here in group theory.

We'll take a similar path through group theory as we did through ring theory. The following is our rough road map for group theory:

(1) Groups and their (basic) properties
(2) Subgroups
(3) Homomorphisms, Isomorphisms
(4) Congruence, Quotients, Homomorphisms (Here we'll find *normal subgroups* vs *ideals*, normal subgroups will arise as kernels)
(5) Isomorphism and Structure Theorems (Another First Isomorphism Theorem, some structure results)

A prototypical example is the group of *permutations*. A permutation on a set T is just a ordering of its elements. Let $T = \{1, 2, 3\}$. There are six possible permutations of T:

$$123 \qquad 132 \qquad 213 \qquad 231 \qquad 312 \qquad 321$$

Each ordering determines a bijective function from T to T: map 1 to the first element of the ordering, 2 to the second, and 3 to the third, i.e. for the ordering 231 there is a bijection $f : T \to T$ given by

$$f(1) = 2 \qquad f(2) = 3 \qquad f(3) = 1$$

Conversely, every bijective function from T to T defines an ordering of elements, namely $f(1)f(2)f(3)$.

A *permutation of a set* T is a bijective function from T to T. Let's go back to our example. Let $T = \{1, 2, 3\}$. It is conventional to represent the permutation $f : T \to T$ whose rule is given by $f(1) = 2$, $f(2) = 3$, $f(3) = 1$ as the array

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$$

We think of the first row as the domain and the second row as the range, i.e. $1 \in T$ gets mapped to $f(1) = 2$.

The composition of two bijective functions is bijective, thus the composition of any two permutations is also a permutation. For instance, take

$$f = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \qquad \text{and} \qquad g = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

[26] Correspondence to Ronald Brown.

Then $f \circ g$ is the function given by

$$f \circ g = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$$

The conventional notation for a permutation let's us easily compute what a composition is by tracking the element through both bijections.

Denote the set of permutations of $T = \{1, 2, 3\}$ by S_3. The composition of functions is an operation on S_3, i.e. $f, g \in S_3$ then $f \circ g \in S_3$. Composition of functions is associative, thus

$$(f \circ g) \circ h = f \circ (g \circ h)$$

There is an identity permutation, which is the identity function, denoted

$$I = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$$

Then $I \circ f = f$ and $f \circ I = I$ for each $f \in S_3$. Furthermore every bijection has an inverse function (which is also a bijection), for instance

$$f = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \qquad f^{-1} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

Thus for $f \in S_3$ then there exists $g \in S_3$ such that $f \circ g = I$ and $g \circ f = I$.

Is $f \circ g = g \circ f$?

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$$

but

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

Definition 7.1. A *group* is a nonempty set G equipped with a binary operation $\cdot$ that satisfies

(1) Closed. If $a, b \in G$ then $a \cdot b \in G$
(2) Associative. $a \cdot (b \cdot c) = (a \cdot b) \cdot c$
(3) Identity. There exists $e \in G$ such that $a \cdot e = a = e \cdot a$
(4) Inverse. For each $a \in G$ there exists $b \in G$ such that $a \cdot b = e = b \cdot a$

A group is called *abelian*[27] if it is commutative, i.e. $a \cdot b = b \cdot a$ for all $a, b \in G$. We will often abbreviate $a \cdot b$ by ab (suppressing the $\cdot$ notation). If G is not abelian it is called *nonabelian*. For abelian groups we will often denote the group operation by $+$, i.e. $a \cdot b$ we'll denote $a + b$ (this is conventional, as we have all seen that addition as we understand it is commutative).

A group G is said to be *finite* or of *finite order* if $|G| < \infty$ (i.e. finite cardinality). In this case $|G|$ is called the *order* of G. Groups with infinitely many elements are said to have *infinite order*.

We've seen that S_3 is a nonabelian group of order 6 with the operation $\cdot$ is $\circ$, i.e. being composition.

The permutation group S_3 is a special case of a general symmetric group. Let n be a positive integer. Let $T = \{1, 2, 3, \ldots, n\}$. Let S_n be the set of all permutations of T (all

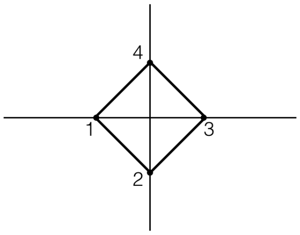
[27] In honor of N. Abel https://en.wikipedia.org/wiki/Niels_Henrik_Abel

bijections $T \to T$). This is called the *symmetric group on* n *symbols*. What is the order of S_n? $|S_n| = n! = n(n-1)(n-2)\cdots(2)(1)$.

7.1. Symmetries of the Square.

> *Il est peu de notions en mathématiques qui soient plus primitives que celle de loi de composition.*
>
> – Nicolas Bourbaki[28]

Let's work an example on the symmetries of the square. This group is denoted D_4 and called the *dihedral group of degree* 4 or the group of symmetries of the square. This is Example 5 on Page 173 in BH. We can consider a 'Platonic' square $\mathbf{S}$ which resides in the plane, whose vertices we identify with the numbers $1-4$.



A *symmetry* of the square is a rigid motion of the square in the plane, i.e. rotating it, flipping it, etc[29]. Mathematically, it is a symmetry is distance-preserving map $m : \mathbf{S} \to \mathbf{S}$. Thus $\mathbf{S} \mapsto m(\mathbf{S})$. These symmetries will become the elements of our group of symmetries. Let's write down a few of them:

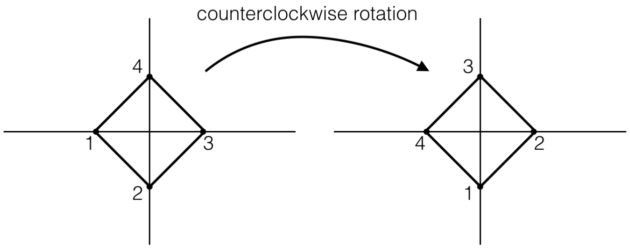


FIGURE 4. Here's a 90° rotation counterclockwise. $\mathbf{S}$ gets mapped to itself, and the image is rotated. Denote this by r

[28] Roughly: *'There are few notions in mathematics that are more primitives than that of law of composition.'* More info on Nicolas Bourbaki: https://en.wikipedia.org/wiki/Nicolas_Bourbaki

[29] The *symmetry* group of a mattress (or rectangle) is a subgroup of the symmetries of a square! See S. Strogatz's NYT article [6]

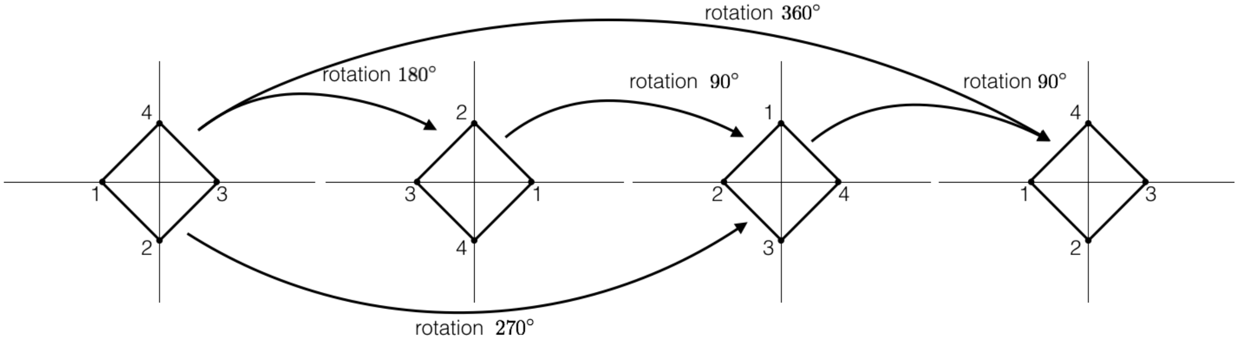


Figure 5. All counterclockwise rotations.

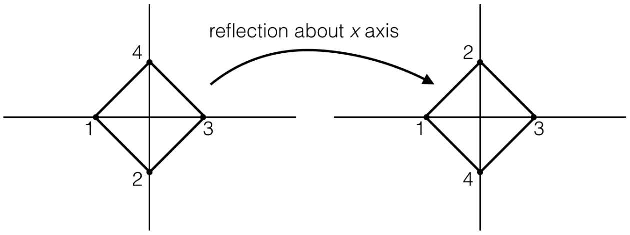


Figure 6. This is a reflection of $\mathbf{S}$ over the x axis. Denote this by s

How many symmetries of the square are there? Any rotation can be written as r^n, i.e. as the n-fold composition $r \circ r \cdots \circ r$ for $n = 0, 1, 2, 3$. How many reflections are there? There are four lines to reflect over, x-axis, y-axis, $x = y$ and $x = -y$. I claim that any of these reflections can be obtained by a composition of s and some rotation. For instance, how to get the reflection over $x = y$?

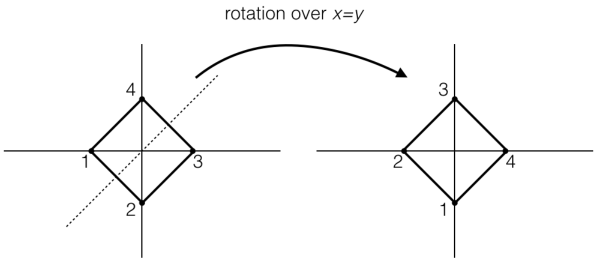


Figure 7. This is a reflection of $\mathbf{S}$ over $x = y$.

There are two other rotations, that may be obtained from r and s as follows:

This means that D_4 is *generated* by r and s, since any element of D_4 can be written as a product of these two elements. The composition operation table is given in BH on page

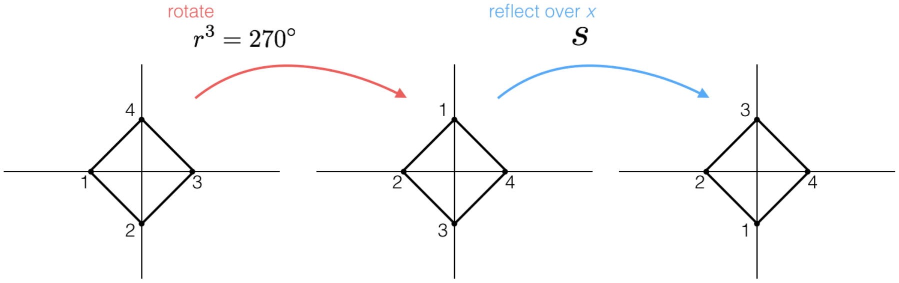


FIGURE 8. First rotate using $r^3 = r \circ r \circ r$. This is a rotation of 270°. Now reflect using s over x-axis.

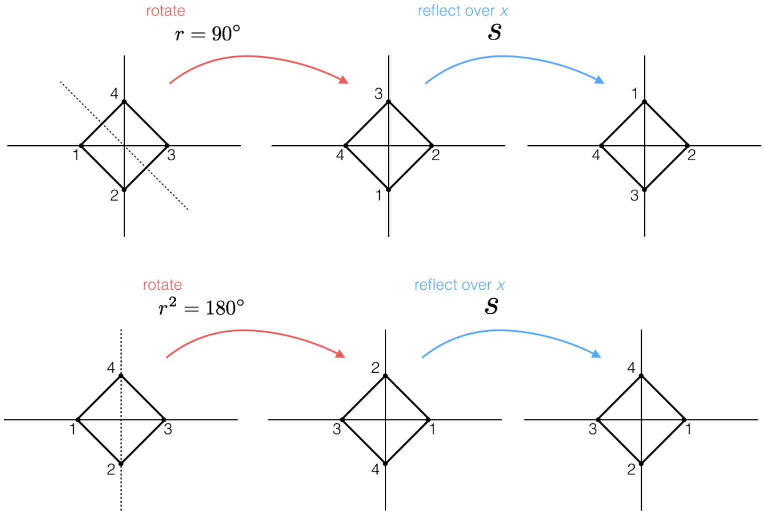


FIGURE 9. Reflections over $y = -x$ and y axes

190. D_4 is not abelian. Why? Consider $s \circ r$ and $r \circ s$? D_4 is a *subgroup* of S_4, the set of permutations on 4 letters. Why?

R. Penrose has an excellent formalization of these symmetries in his wonderful book [5, Chapter 13.1]. We will follow his exposition. Let's think about the Platonic square **S** as residing in the complex plane $\mathbb{C}$. We may represent the vertices of the square as the points $1, i, -1, -i$ as in Figure 10.

In Figure 10 we see the (subgroup of) rotations can be represented by multiplication by $1 = i^0, i, -1 = i^2, -i = i^3$. Our reflection s above corresponds to the function $C : \mathbb{C} \to \mathbb{C}$, complex conjugation $z \xrightarrow{C} \bar{z}$, i.e. reflection over the x-axis. Thus the reflections are C, iC, i^2C, i^3C.

7.2. **Groups and Rings.** We know a ring R has two associative operations. When is R is a group under one?

Proposition 7.2. *Let* R *be a ring. Then* $(R, +)$ *is an abelian group.*

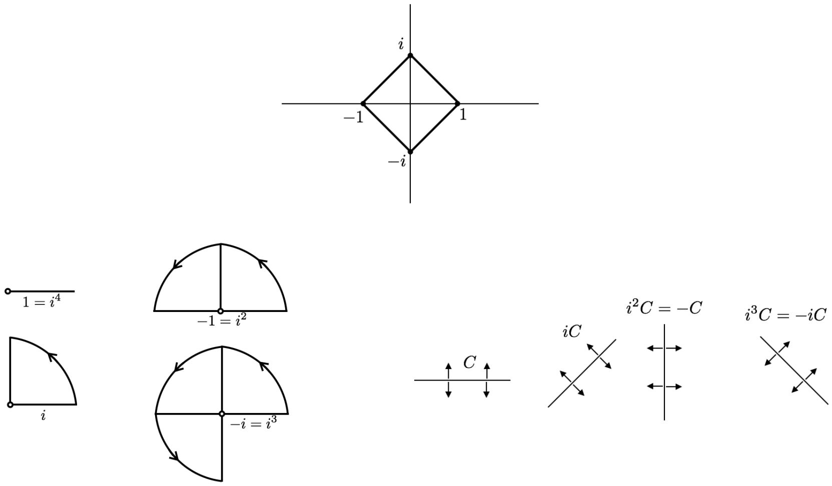


Figure 10. The square embedded in $\mathbb{C}$ with its reflections and rotations.

Proof.

1. Closed (?) Yes.
2. Associative (?) Yes.
3. Identity (?) Yes.
4. Inverses (?) Yes.

□

Thus the following familiar rings are abelian groups under addition

$$\mathbb{Z}, \mathbb{Z}_n, \mathbb{Q}, \mathbb{R}, \mathbb{C}, M_2(\mathbb{R}), R[x]$$

Conjecture 7.3. *Let R be a nonzero ring. Is $(R, \cdot)$ a group?*

Proof.

1. Closed (?) Yes.
2. Associative (?) Yes.
3. Identity ? Not necessarily.
4. Inverse: 0_R has no inverse. Axiom 4 always fails.

□

Thus for a nonzero ring R, $(R, \cdot)$ is never a group under multiplication. If R does not have identity it fails the identity axiom. Even if does have identity, then 0_R has no inverse and the inverse axiom fails. How can we fix this?

Proposition 7.4. *Let F be a field. Let F^* be the nonzero elements of F. Then F^* is an abelian group under multiplication.*

Proof. From the ring axioms we have $1-4$. From commutativity we have 5.

1. Closed.
2. Associative.
3. Identity.
4. Inverse.
5. Abelian.

□

The notation F^* is conventional the $*$ refers to multiplication. Sometimes you'll see this as $F^\times$. For instance, $\mathbb{Q}^*, \mathbb{R}^*, \mathbb{C}^*$ are all abelian groups under multiplication.

Non-example. Is $\mathbb{Z}\backslash\{0\}$ a group under multiplication? (Inverses?) What about $\mathbb{Z}_{10}\backslash\{0\}$ with multiplication? (Again inverses?) How about $\mathbb{Z}_n$ for n composite?

Theorem 7.5. *Let R be a ring with identity. The set U of all units in R is a group under multiplication.*

Proof.
(1) U is closed as we've seen the product of two units is a unit. Let $a, b \in U$. Then $(ab)(b^{-1}a^{-1}) = 1 = (b^{-1}a^{-1})ab$.
(2) Associative follows from ring axiom.
(3) 1_R is a unit, thus $1_R \in U$
(4) U has inverses by definition of unit.

□

Example. Denote the group of units in $\mathbb{Z}_n$ by U_n. Then this is a group. What is U_n? (Think about coprime).

Example. Consider

$$GL_2(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{R} \quad ad - bc \neq 0 \right\}$$

$GL_2(\mathbb{R})$ is called the *general linear group* (the set of invertible matrices). It is an infinite order nonabelian group.

Just as with rings, we can consider a *product group*.

Theorem 7.6. *Let G and H be groups. Define $G \times H$ by*

$$(g, h)(g', h') = (gg', hh')$$

Then $G \times H$ is a group. If G and H are abelian then so is $G \times H$. If G and H are finite, so is $G \times H$ and $|G \times H| = |G||H|$.

For instance, consider $(\mathbb{Z}, +)$ and $(\mathbb{Z}_6, +)$ with addition. We can make a product group $\mathbb{Z} \times \mathbb{Z}_6$. What is the identity - $(0, 0)$. What is the inverse of $(7, 4)$? How about $(-7, 2)$, we have $(7, 4) + (-7, 2) = (0, 0)$.

7.3. Basic Properties.

In order to speak of *the* inverse element and *the* identity element we must show these objects are unique.

Theorem 7.7. *Let G be a group and let $a, b, c \in G$. Then*
(1) *G has a unique identity element*
(2) *Cancellation holds: If $ab = ac$ then $b = c$ and if $ba = ca$ then $b = c$.*
(3) *Each element of G has a unique inverse.*

Proof.
(1) Suppose e, e' are identities. Then $e = ee' = e'$.
(2) If $ab = ac$ then $a^{-1}(ab) = a^{-1}(ac)$ implying $b = c$.
(3) Let $a \in G$. Let b, c be inverses of a. Then $ab = e = ac$. By cancellation $b = c$.

□

Because the inverse is unique, we denote an inverse as a^{-1}.

Corollary 7.8. *Let G be a group. Let $a, b \in G$. Then*
(1) $(ab)^{-1} = b^{-1}a^{-1}$

(2) $(a^{-1})^{-1} = a$

Proof. We see that $(ab)(b^{-1}a^{-1}) = 1 = (b^{-1}a^{-1})(ab)$. We have $aa^{-1} = e = a^{-1}a$. Thus a is an inverse for a^{-1}. By uniqueness $(a^{-1})^{-1} = a$. □

Let G be a group and $a \in G$. We define $a^n = aa a \cdots a$ (n factors). We define $a^0 = e$. And $a^{-n} = a^{-1} \cdots a^{-1}$ (n factors).

An element $a \in G$ is said to have *finite order* if $a^k = e$ for some $k > 0$. The *order of* a is said to be the smallest positive integer n such that $a^n = e$. The order of a is denoted $|a|$. An element has *infinite order* if $a^k \neq e$ for all $k > 0$.

Order of 8 in $\mathbb{Z}_{12}$? (See $8 + 8 + 8 = 24$).

Theorem 7.9. *Let* G *be a group and* $a \in G$*.*

(1) *If* a *has infinite order, then the* a^k *are all distinct*
(2) *If* $a^i = a^j$ *for* $i \neq j$ *then* a *has finite order.*

Proof. Notice that these statements are contrapositives. Thus it suffices to prove one of them. Suppose $a^i = a^j$ for $i > j$. Multiply both sides by a^{-j} to get $a^{i-j} = a^{j-j} = a^0 = e$. Since $i - j > 0$ this implies a has finite order. □

Theorem 7.10. *Let* G *be a group and* $a \in G$ *with* $|a| = n$*. Then*

(1) $a^k = e$ *if and only if* $n|k$
(2) $a^i = a^j$ *if and only if* $i \equiv j \mod n$
(3) *If* $n = td$ *with* $d \geqslant 1$ *then* a^t *has order* d

Proof.
(1) If $n|k$ say $k = nt$. Then $a^k = a^{nt} = (a^n)^t = e^t = e$. Now suppose $a^k = e$. Use the division algorithm to write $k = nq + r$ with $0 \leqslant r < n$. Thus $e = a^k = a^{nq+r} = a^{nq}a^r = (a^n)^q a^r = e^q a^r = ea^r = a^r$. By definition of order n is smallest positive integer with $a^n = e$. Thus $a^r = e$ implies $r = 0$ and $n|k$.
(2) $a^i = a^j$ if and only if $a^{i-j} = e$. By (1) we have $a^{i-j} = e$ if and only if $n|(i-j)$ thus if and only if $i \equiv j \mod n$.
(3) We have $e = a^n = a^{td} = (a^t)^d$. We must show d is smallest such integer. If $(a^t)^k = e$ for some positive integer k then $a^{tk} = e$. Thus $n|tk$ by (1) so $tk = mn = m(td)$. Thus $k = md$, so $d \leqslant k$.

□

8. Subgroups

Definition 8.1. Let G be a group. $H \subseteq G$ is a *subgroup* if H is itself a group. This is denoted $H \leqslant G$.

Every group G has two subgroups: $G \leqslant G$ and $\{e\} \leqslant G$. $\{e\}$ is called the *trivial subgroup*.

We've seen that $\mathbb{R}^*$ - the nonzero real numbers is a group under multiplication. The group $\mathbb{R}^{**}$ of positive real numbers is a subgroup of $\mathbb{R}^*$. $\mathbb{Z} \subset \mathbb{Q}$ is a subgroup.

When checking that a subset is a subgroup, we only have to check two of the axioms:

Proposition 8.2. *A nonempty subset* $H \subset G$ *is a subgroup provided that*

(1) *If* $a, b \in H$ *then* $ab \in H$ *(closure)*

(2) *If* $a \in H$ *then* $a^{-1} \in H$ *(inverses)*

Proof. Associativity holds since $H \subseteq G$. By hypothesis we have inverses and closure. It remains to show $e \in H$. H is nonempty so there is some $a \in H$. By (2) we have $a^{-1} \in H$. Thus $e = aa^{-1} \in H$. □

Consider

$$H := \left\{ \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} : x \in \mathbb{R} \right\}$$

We have $1 \cdot 1 - x \cdot 0 = 1$ so H is a nonempty subset of $GL_2(\mathbb{R})$. We have $\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & y \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & x+y \\ 0 & 1 \end{pmatrix}$. The inverse of $\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$ is $\begin{pmatrix} 1 & -x \\ 0 & 1 \end{pmatrix}$. Thus H is a subgroup by Proposition 8.2.

Proposition 8.3. *Let* H *be a nonempty finite subset of a group* G. *If* H *is closed under the operation in* G, *then* H *is a subgroup of* G.

Proof. By Proposition 8.2 we need only verify that inverses are in H. If $a \in H$, by closure $a^k \in H$ for every $k \geqslant 1$. H is finite, so $a^j = a^i$ for some $i \neq j$. Thus $|a| = n < \infty$. Now $(a^{n-1})a = a^n = e = a^n = a(a^{n-1})$. So a^{n-1} is the inverse of a and is in H. □

Consider $H = \{\sigma \in S_5 : \sigma(1) = 1\}$, i.e. all permutations on 5 letters that fix 1. If $g, h \in H$ then $g(1) = 1 = h(1)$. So $(g \circ h)(1) = g(h(1)) = 1$. Thus $g \circ h \in H$ and $H \leqslant S_5$ by Proposition 8.3.

Let G be a group. The *center* of G is the subset $Z(G)$ defined as

$$Z(G) := \{a \in G : ag = ga \text{ for all } g \in G\}$$

Thus $Z(G)$ is the set of elements that commute with every element of G. Notice that if G is abelian then $Z(G) = G$. When G is nonabelian then $Z(G) \neq G$. $Z(G)$ is always nonempty, as $e \in Z(G)$.

Theorem 8.4. *Let* G *be a group. Then* $Z(G) \leqslant G$.

Proof. $Z(G)$ is nonempty. By Theorem 8.2 we must show $Z(G)$ is closed and has inverses. Let $a, b \in Z(G)$. Let $g \in G$. Then

$$(ab)g = a(bg) = a(gb) = (ag)b = (ga)b = g(ab)$$

We have $ga^{-1} = (ag^{-1})^{-1} = (g^{-1}a)^{-1} = a^{-1}g$. □

An important type of subgroup is constructed from a single element. Let G be a group. Let $a \in G$. Then

$$\langle a \rangle = \{\ldots, a^{-3}, a^{-2}, a^{-1}, a^0, a^1, a^2, a^3, \ldots\} = \{a^n | a \in \mathbb{Z}\}$$

Proposition 8.5. *Let* G *be a group. Let* $a \in G$. *Then* $\langle a \rangle \leqslant G$.

Proof. $a \in \langle a \rangle$. $a^i a^j = a^{i+j} \in \langle a \rangle$. The inverse of a^k is $a^{-k} \in \langle a \rangle$. □

Theorem 8.6. *Let* G *be a group. Let* $a \in G$.

(1) *If* a *has infinite order, then* $\langle a\rangle$ *is an infinite subgroup consisting of the distinct elements* a^k *with* $k \in \mathbb{Z}$

(2) *If* $|a| = n$ *then* $|\langle a\rangle| = n$ *and* $\langle a\rangle = \{e, a^1, a^2, \dots a^{n-1}\}$.

Proof. (1) The follows directly from our previous lemma.

(2) Let $a^i \in \langle a\rangle$. Then $i \cong k \mod n$ for some $k \in \{0, 1, 2, \dots n-1\}$. Thus $a^i = a^k$. Furthermore all powers a^k for $k \in \{0, 1, 2, \dots n-1\}$ are distinct.

□

Let G be a group. Let $S \subseteq G$ be nonempty. Define $\langle S\rangle$ as the set of all possible products of elements of S and their inverses.

Theorem 8.7. (1) $\langle S\rangle \leqslant G$ *and* $S \subset \langle S\rangle$.
(2) *If* $H \leqslant G$ *and* $S \subseteq H$ *then* $\langle S\rangle \subseteq H$.

Proof. (1) $\langle S\rangle$ is nonempty as S is nonempty. Let $a, b \in \langle S\rangle$. Then $a = a_1a_2\cdots a_k$ where $a_i \in S$ or $a_i^{-1} \in S$ and $b = b_1b_2\cdots b_m$ where $b_i \in S$ or $b_i^{-1} \in S$. Thus $ab = a_1a_2\cdots a_kb_1b_2\cdots b_m$ consists of elements of S or their inverses. Hence $ab \in \langle S\rangle$. The inverse of $a = a_1a_2\cdots a_k \in S$ is $a_1^{-1}a_2^{-1}\cdots a_k^{-1}$ and a_i^{-1} is either an element of S or inverse of element in S since a_i is. Thus $a^{-1} \in S$. Thus $\langle S\rangle \leqslant G$.

(2) Any subgroup H with $S \subset H$ must have all inverses of the elements of S. By closure it must also contain all possible products of elements of S and their inverses. Thus $\langle S\rangle \subseteq H$.

□

This theorem shows that $\langle S\rangle$ is the smallest subgroup of G that contains S (this is an alternate definition). $\langle S\rangle$ is called the *subgroup generated by* S.

9. Homomorphisms and Isomorphisms

We've seen the idea of homomorphism and isomorphism before. In this Section we'll write down what the morphisms are in the category of groups.

Definition 9.1. *Let* G, H *be groups.* G *is* isomorphic *to* H *if there is a function* $f : G \to H$ *such that*

(1) f *is injective*
(2) f *is surjective*
(3) $f(ab) = f(a)f(b)$ *for all* $a, b \in G$

If G and H are isomorphic we denote this by $G \cong H$. The groups isomorphic to a given group G form the *isomorphism class* of G and any two groups in that class are isomorphic. Sometimes you hear about *classifying groups* and what is meant by this is describing the isomorphism classes.

Consider the subset of $GL_2(\mathbb{R})$ given by

$$H := \left\{ \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} : x \in \mathbb{R} \right\}$$

Let $A = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & y \\ 0 & 1 \end{pmatrix}$. Then $AB = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & y \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & x+y \\ 0 & 1 \end{pmatrix}$

The upper right entries add when matrices multiply in H. The rest of the entries remain fixed. Thus when computing with such matrices we only need to keep track of the upper right entry. So what does H look like? It seems to look like $\mathbb{R}$. How about the map $a \mapsto \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$. Exercise: show this is an isomorphism.

The additive group $\mathbb{R}$ is isomorphic to $\mathbb{R}^{**}$ of positive real numbers. Let $f : \mathbb{R} \to \mathbb{R}^{**}$ be given by $a \mapsto 10^a$.

(1) (Injective) If $10^r = 10^s$ then take log to get $\log 10^r = \log 10^s$ implying $r = s$.
(2) (Surjective.) Let $r \in \mathbb{R}$. Then $a = \log r$. Then $f(a) = 10^a = 10^{\log r} = r$.
(3) (Homomorphism.) Finally $f(a + b) = 10^{a+b} = 10^a 10^b = f(a)f(b)$

Some remarks on isomorphisms.

(1) If G and H have different orders, then G and H are not isomorphic.
(2) If G is abelian and H is nonabelian then G and H are not isomorphic (compare $\mathbb{Z}_6$ and S_3).[30]
(3) It is straightforward that if f is an isomorphism then a and $f(a)$ have the same order. Let's do an example. Are $\mathbb{Z}_2 \times \mathbb{Z}_2$ and $\mathbb{Z}_4$ isomorphic? These are both abelian groups, with the same order. However, $\mathbb{Z}_4$ has *elements* of order 4, namely 1 and 3. $\mathbb{Z}_2 \times \mathbb{Z}_2$ only has elements of order 1 and 2.

If G is a group then an isomorphism $G \to G$ is called an *automorphism* of G. A simple example of an automorphism is the identity map $\iota : G \to G$ given by $G \ni g \mapsto g \in G$. It's straightforward that this is indeed an automorphism. Here is an important example of an automorphism. Let G be a group. Fix $a \in G$. Define $f : G \to G$ by $f(g) = a^{-1}ga$. f is sometimes called *conjugation by* a[31]. Then

$$f(gh) = a^{-1}gha = a^{-1}g(aa^{-1})ha = (a^{-1}ga)(a^{-1}ha) = f(g)f(h)$$

Now we show that f is injective. $\ker f = \{g \in G : a^{-1}ga = e_G\} = \{e_G\}$. Let's show f surjective. Let $g \in G$. Consider $h = aga^{-1}$. Then $f(h) = a^{-1}ha = a^{-1}(aga^{-1}a = (a^{-1}a)g(a^{-1}a) = ege = g$. Thus $g \mapsto a^{-1}ga$ is an automorphism. This is sometimes called the *inner automorphism* of G by a.

9.1. **Homomorphisms.**

Definition 9.2. Let G, H be groups. A function $f : G \to H$ is a *homomorphism* if

$$f(ab) = f(a)f(b) \quad \text{for all } a, b \in G$$

There are important subgroups associated to every homomorphism. These are the kernel and image. The *kernel* of a group homomorphism $f : G \to H$ is defined as $\ker f := f^{-1}(e_H) = \{g \in G : f(g) = 0\}$.

Consider $f : \mathbb{R}^* \to \mathbb{R}^*$ given by $f(x) = x^2$. This is a homomorphism as $f(xy) = (xy)^2 = x^2y^2 = f(x)f(y)$. However, f is not injective as $(-x)^2 = f(-x) = f(x) = x^2$.

Theorem 9.3. *Let* G, H *be groups. Denote the identities* e_G *and* e_H*. For a homomorphism* $f : G \to H$ *we have*

(1) $f(e_G) = e_H$

[30] We'll develop the notion of $\ker f$ in a bit. This will give us more algebraic machinery. In this case, we can examine where $ab - ba$ maps under $f : G \to H$. If G is abelian then $f(ab - ba) = 0$.

[31] Conjugation is an idea in math that arises often. See https://en.wikipedia.org/wiki/Matrix_similarity (do you see the similarity here? no pun intended) and https://en.wikipedia.org/wiki/Conjugacy_class

(2) $f(a^{-1}) = f(a)^{-1}$ *for all* $a \in G$
(3) $\mathrm{Im}(f) \leqslant H$.
(4) *If* f *is injective then* $G \cong \mathrm{Im} f$.

Proof. (1) We have $f(e_G) = f(e_G e_G) = f(e_G)f(e_G)$ Use cancellation, applying $f(e_G)^{-1}$, to get $e_H = f(e_G)$.
(2) $f(a)f(a)^{-1} = e_H = f(e_G) = f(aa^{-1}) = f(a)f(a^{-1})$. By cancellation $f(a^{-1}) = f(a)^{-1}$.
(3) Let $a, b \in \mathrm{Im}(f)$. Then $f(g) = a$ and $f(h) = b$ for some $g, h \in H$. Thus $ab = f(g)f(h) = f(gh) \in \mathrm{Im}(f)$. For inverses we have $f(g^{-1}) = f(g)^{-1} = a^{-1}$. Thus $a^{-1} \in \mathrm{Im}(f)$.
(4) We can consider $f : G \to H$ as a surjective homomorphism $f : G \to \mathrm{Im}(f) \subseteq H$. If f is injective, then f is an isomorphism by definition.

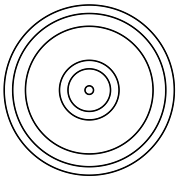
□

Lemma 9.4. *Let* $f : G \to H$ *be a homomorphism. Then* $\ker f = \{e\}$ *if and only if* f *is injective.*

Proof. If f is injective and $g \in \ker f$ then $f(g) = f(e_G)$. Thus $g = e_G$ by injectivity. If $\ker f = \{e\}$ observe that $f(a) = f(b)$ if and only if $f(a)f(b)^{-1} = e_H$ if and only if $f(a)f(b^{-1}) = e$ if and only if $f(ab^{-1}) = e$. Thus $ab^{-1} = e$, so $a = b$. □

Here are some examples of homomorphisms:

(1) the determinant function, $\det : GL_2(\mathbb{R}) \to \mathbb{R}$ given by $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto ad - bc$
(2) Let G be a group. Let $a \in G$. The map $\varphi : \mathbb{Z} \to G$ given $n \mapsto a^n$
(3) Let $H \leqslant G$. The *inclusion map* $i : H \to G$ given by $i(a) = a$.
(4) Consider $\mathbb{C}^*$ and $\mathbb{R}^*$. Let $\varphi : \mathbb{C}^* \to \mathbb{R}^*$ given by $\varphi(a) = |a|$. The fibers of this map are concentric circles about 0 in $\mathbb{C}$, visualized as follows:



(5) $f : \mathbb{Z} \to \mathbb{Z}_5$ given by $f(a) = [a]_5$. $\ker f = 5\mathbb{Z}$.
(6) Let G, G' be groups. The product group $G \times G'$ is related to the factors G and G' by homomorphisms:

$$\begin{tikzcd}
G \arrow[dr, "i"] & & G \\
& G \times G' \arrow[ur, "p"] \arrow[dr, "p'"'] & \\
G' \arrow[ur, "i'"'] & & G'
\end{tikzcd}$$

here $i(x) = (x, e_{G'})$, $\quad i'(x) = (e_G, x)$, $\quad p(x, x') = x \quad p'(x, x') = x'$

It is straightforward i, i', p, p' are homomorphisms. The maps i, i' are injective and p, p' are surjective. $\mathrm{Im}(i) = G \times e_{G'}$ and $\ker p = e_G \times G'$. Similar for i', p'.

9.2. **Cayley's Theorem.** The next theorem relates arbitrary groups with groups of permutations. It says that the study of group theory can be reduced to the study of permutation groups.

Theorem 9.5 (Cayley's Theorem)**.** *Every group* G *is isomorphic to a group of permutations.*

Proof. Consider the group $A(G)$ of all permutations on the *set* G (that is, forget the group structure). $A(G)$ consists of all bijection *functions* from G to G with composition as the group operation. We'll find a subgroup of $A(G)$ isomorphic to G.

Let $a \in G$. we claim that the map $\varphi_a : G \to G$ defined by $\varphi_a(x) = ax$ is a bijection. This follows from the fact that if $ag = ah$ then $g = h$ by cancellation. Thus φ_a is injective. For $h \in G$ we have $a(a^{-1}h) = h$. So φ_a is surjective. Thus $\varphi_a \in A(G)$. What we've shown here is a correspondence between elements $a \in G$ and functions $\varphi_a : G \to G$.

Define $f : G \to A(G)$ by

$$a \mapsto \varphi_a$$

For $a, b \in G$ we have $f(ab) = \varphi_{ab}$. This is the function $\varphi_{ab}(x) = (ab)x = abx$. Observe that

$$\varphi_{ab}(x) = (ab)x = \varphi_a(bx) = \varphi_a\varphi_b(x) = (\varphi_a \circ \varphi_b)(x)$$

This implies that f is in fact a homomorphism. Notice that

$$\ker f = \{a \in G : \varphi_a(x) = e_G\} = \{a \in G : ax = e_G \text{ for all } x \in G\} = \{e_G\}$$

Thus f is injective. This implies by Theorem 9.3 (4) we have $G \cong \operatorname{Im}(f) \leqslant A(G)$.

□

Corollary 9.6. *Let* G *be a group. Let* $|G| = n$. *Then* G *is isomorphic to a subgroup of* S_n.

See the discussion in BH on page 222. These remarks put Cayley's Theorem into context and do well explaining why Cayley's theorem, though powerful, is not too profitable in terms of applicability.

> *In these days the angel of topology and the devil of abstract algebra fight for the soul of every individual discipline of mathematics.*
>
> – Herman Weyl[32]

In this section we'll study congruence in groups. Recall that we've seen this in the $\mathbb{Z}, F[x]$ and general rings R/I. The results will be very similar for groups.

For $a \equiv b \mod 4$ means that $4|a-b$, i.e. $a-b$ is a multiple of 4. Let

$$K = \{0, \pm 4, \pm 8, \ldots\}$$

Thus

$$a \equiv b \mod 4 \iff a - b \in K$$

Notice that $K \leqslant \mathbb{Z}$. What subgroup is it? $K = \langle 4 \rangle$, the cyclic subgroup generated by 4. Thus instead of thinking of congruence modulo an element 4, we think of congruence modulo the subgroup K:

$$a \equiv b \mod K \iff a - b \in K$$

We can apply this idea for any subgroup $K \leqslant G$, not just the cyclic subgroups. This will motivate the definition of congruence for groups - Definition 8.1; recall that in general group notation in multiplicative.

To paraphrase David Foster Wallace, we may lodge the obvious complaint/observation here: 'We've Seen This Before'.[33] This should all look quite familiar by now:

(1) We introduce a definition and notation for congruence $a \equiv b \mod K \iff a - b \in K$
(2) We show thinking about congruence is equivalent to considering cosets, i.e. $Ka = Kb \iff a = b \mod K$
(3) We introduce a (quotient group) operation on cosets $(Ka) \cdot (Kb) = K(ab)$, we show that K must be a *normal* subgroup for this operation to be well defined (this corresponds to ideals in rings)
(4) We show normal subgroups arise as kernels, and relate homomorphisms and quotient groups

8. Group Theory: Congruence, Normality and Quotients

For rings we defined congruence using ideals I. We said that $a \equiv b \mod I$ if $a - b \in I$. We now frame this for groups.

Definition 8.1. Let $K \leqslant G$. Let $a, b \in G$. Then a *is congruent to* b modulo K, written $a \equiv b \mod K$, if $ab^{-1} \in K$.

Theorem 8.2. *Let $K \leqslant G$. Then the relation of congruence modulo K is*

(1) *reflexive:* $a \equiv a \mod K$
(2) *symmetric: if* $a \equiv b \mod K$ *then* $b \equiv a \mod K$
(3) *transitive: if* $a \equiv b \mod K$ *and* $b \equiv c \mod K$ *then* $a \equiv c \mod K$

[32] *Weyl, H. (1939). Invariants. Duke Mathematical Journal, 5(3), 489-502.* Note that this was about 80 years ago. By now algebra and topology have joined forces to form algebraic topology; any dueling for souls is perhaps now being done against hedge funds.

[33] *DFW. Consider the Lobster.*

Proof. (1) $aa^{-1} \in K$. Thus $a \equiv a \mod K$.

(2) Let $a \equiv b \mod K$. Then $ab^{-1} \in K$. Since K is a subgroup $ba^{-1} = (ab^{-1})^{-1} \in K$.

(3) Let $a \equiv b \mod K$ and $b \equiv c \mod K$. Then $ac^{-1} = a(b^{-1}b)c^{-1} = (ab^{-1})(bc^{-1}) \in K$.

□

Let $K \leqslant G$. Let $a \in G$. The *congruence class* of a modulo K is the set of elements of G congruent to a modulo K, i.e.

$$\{b \in G : b \equiv a \mod K\} = \{b \in G : ba^{-1} = k \in K\} = \{ka : k \in K\}$$

We define $Ka = \{ka : k \in K\}$. This is called a *right coset* of K in G. When the group is abelian, i.e. the operation is denoted $+$, then we use the notation $a + K$ for the coset.[34]

Proposition 8.3. *Let $K \leqslant G$. Let $a, c \in G$. Then $a \equiv c \mod K$ if and only if $Ka = Kc$.*

Proof. Let $a \equiv c \mod K$. Let $x \in Ka$. Then $x = ka$. We have $ac^{-1} = k' \in K$. Thus $x = ka = k(k'c) = (kk')c$. So $x \in Kc$ and $Ka \subseteq Kc$. By symmetry $c = a \mod K$ and a similar argument shows $Ka \subseteq Kc$.

Let $Ka = Kc$. $a = ea \in Ka = Kc$. Thus $a = kc$ so $ac^{-1} \in K$. □

Corollary 8.4. *Let $K \leqslant G$. Then either $Ka \cap Kc = \emptyset$ or $Ka = Kc$.*

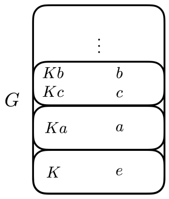


FIGURE 11. A schematic diagram for cosets of $K \leqslant G$. Notice that $e \in K$ since K is a subgroup. $Kb = Kc \iff b \equiv c \mod K \iff b - cK$. Another way to think about this is that we are setting/collapsing the entire subgroup K 'to zero'. Thus '$b = c$' $\iff b - c \in K$ (i.e. versus $b = c \iff b - c = 0$). This is what it means to think modulo K.

We'll develop some facts about cosets.

Theorem 8.5. *Let $K \leqslant G$. Then*

(1) *G is the union of right cosets of K: $G = \bigcup_{a \in G} Ka$*

(2) *For each $a \in G$ there is a bijection $f : K \to Ka$. Thus if K is finite, any two right cosets have the same number of elements*

Proof. (1) Every right coset $Ka \subseteq G$. Thus $\bigcup_{a \in G} Ka \subseteq G$. Let $b \in G$. Then $b \in Kb \subseteq \bigcup_{a \in G} Ka$.

(2) Define $f : K \to Ka$ by $f(x) = xa$. Then if $f(x) = f(y)$ we have $xa = ya$ implying $x = y$. So f is injective. For $b \in Ka$ we have $b = ka$ for $k \in K$. Thus $f(k) = ka = b$. Thus f is surjective. (Is f a homomorphism? No, $f(ab) = kab \neq kakb$ in general).

□

[34] This should look very familiar - this is precisely what we did for rings.

If $H \leqslant G$ the number of distinct right cosets of H in G is called the *index of* H *in* G. This is denoted $[G : H]$. If G is finite, there are only a finite number of cosets so $[G : H]$ is finite. If G is infinite, then $[G : H]$ may be infinite or finite.

Here's an example. Consider the group $\mathbb{Z}$. Let $H = \langle 3 \rangle$. Recall $\langle 3 \rangle = 3\mathbb{Z}$. The cosets of H are the congruence classes mod H. We have $H, 1+H, 2+H$. Thus $[\mathbb{Z} : H] = 3$.

Any example of infinite index would be, say, $[\mathbb{Q} : \mathbb{Z}]$. This holds, for instance, as the numbers $\frac{1}{n}$ for $n > 1$ have distinct cosets relative to $\mathbb{Z}$.

Theorem 8.6 (Lagrange's Theorem). *Let* $K \leqslant G$ *with* $|G| < \infty$. *Then* $|G| = [G : K]|K|$. *In particular* $|K| \mid |G|$, *i.e. the order of* K *divides the order of* G.

Proof. Suppose $[G : K] = n$. Then there are n distinct cosets $Kc_1, Kc_2, \ldots Kc_n$. We know

$$G = Kc_1 \cup Kc_2 \ldots \cup Kc_n$$

by Theorem 8.5. These cosets are distinct, thus mutually disjoint by Corollary 8.4. Therefore $|G| = |Kc_1| + |Kc_2| + \ldots + |Kc_n|$.[35] For each c_i we have $|Kc_i| = |K|$ by Theorem 8.5. Thus

$$|G| = n|K| = |K|[G : K]$$

□

Corollary 8.7. *Let* G *be a group with* $|G| < \infty$.

(1) *If* $a \in G$ *then the order of* $|a|$ *divides* $|G|$
(2) *If* $|G| = k$ *then* $a^k = e$ *for every* $a \in G$

Proof. (1) We have that $|a| = |\langle a \rangle|$. Thus $|G| = [G : \langle a \rangle]|a|$.
(2) If $|a| = n$. Then $n|k$. Thus $k = nt$ so $a^k = a^{nt} = (a^n)^t = e^t = e$.

□

8.1. **The Structure of Finite Groups.** See Page 242 in BH for remarks on the classification of groups. We'll recapitulate these remarks.

A major goal of group theory is to classify all finite groups up to isomorphism, i.e. produce a list of groups such that every finite group is isomorphic to exactly one group on the list. This is basically a look up table for isomorphism. This is very difficult in general.[36] We'll see in this section that we can classify all cyclic groups and groups of prime order (which turn out to just be cyclic groups). We'll do this using Lagrange's theorem, and the idea that the order of elements and subgroups must divide the order of the group. We'll also see some classification for groups of small order.

Here is a classification of cyclic groups.

Theorem 8.8 (Theorem 7.19 in BH). *Let* G *be a cyclic group.*

(1) *If* G *is infinite, then* $G \cong \mathbb{Z}$.
(2) *If* $|G| = n$ *then* $G \cong \mathbb{Z}_n$

[35] Notice this follows from an elementary set theoretic fact about cardinalities: if $A \cap B = \emptyset$ then $|A \cup B| = |A| + |B|$.

[36] See Chapter 9 in BH for more techniques to do this: e.g. Sylow theorems, structure theorem for finite abelian groups, etc.

Proof.

(1) Let $G = \langle a \rangle = \{a^k : k \in \mathbb{Z}\}$. Define $f : G \to \mathbb{Z}$ by $f(a^k) = k$. G infinite implies $a^i = a^j$ if and only if $i = j$ (Theorem 7.15), thus f is injective. f is surjective as for $k \in \mathbb{Z}$ we have $a^k \mapsto k$. $f(a^i a^j) = f(a^{i+j}) = i + j = f(a^i) + f(a^j)$. Thus $G \cong \mathbb{Z}$.

(2) Let $G = \langle b \rangle$ with $|b| = n$. $G = \{b^0, b^1, \dots, b^{n-1}\}$ by Theorem 7.15. We have $\mathbb{Z}_n = \{[0], [1], \dots, [n-1]\}$. Define $g : G \to \mathbb{Z}_n$ by $g(b^i) = [i]$. g is clearly a bijection and $g(b^i b^j) = g(b^{i+j}) = [i+j] = [i] + [j] = g(b^i) + g(b^j)$. Thus $G \cong \mathbb{Z}_n$.

□

Here is a classification for groups of prime order.

Theorem 8.9. *Let $|G| = p$ with p prime. Then G is cyclic and $G \cong \mathbb{Z}_p$.*

Proof. Let $a \in G$ with $a \neq e$. Then $|a| | p$. But $|a| > 1$. Thus $|a| = p$. Therefore $\langle a \rangle = G$ and G is a cyclic group of order p, so $G \cong \mathbb{Z}_p$ by Theorem 9.7. □

Now here are some classifications for groups of small order.

Theorem 8.10. *Let G be a group with $|G| = 4$. Then $G \cong \mathbb{Z}_4$ or $G \cong \mathbb{Z}_2 \times \mathbb{Z}_2$.*

Proof. If G contains an $a \in G$ of order 4 then $\langle a \rangle = G$ so $G \cong \mathbb{Z}_4$. If G has no elements of order 4, then for $a \in G$ with $a \neq e$ we have $|a| = 2$. Thus $G = \{e, a, b, c\}$ with $a = a^{-1}, b = b^{-1}, c = c^{-1}$. Thus the product any two distinct a, b, c is the third, i.e. $ab = ba = c, cb = bc = a, ca = ac = b$. Define $f : G \to \mathbb{Z}_2 \times \mathbb{Z}_2$ as

$$e \mapsto (0,0) \qquad a \mapsto (1,0) \qquad b \mapsto (1,0) \qquad c \mapsto (1,1)$$

Then, for instance, $f(ac) = f(b) = (1,0) = (0,1) + (1,1) = f(a) + f(c)$. It's straightforward that this holds in general. □

The following theorem appears as Theorem 8.9 in BH. You should read this proof as it is done with very elementary group theory. For our part, we'll come back and proof this once we have more algebraic tools.[37]

Theorem 8.11. *Let G be a group with $|G| = 6$. Then $G \cong \mathbb{Z}_6$ or $G \cong S_3$.*

As summarized in BH, page 245, we now have classification for all groups of order less than 7. For $|G| = 2, 3, 5, 7$ we have that these are primes, thus $G \cong \mathbb{Z}_p$. For $|G| = 4$ we showed $G \cong \mathbb{Z}_4, \mathbb{Z}_2 \times \mathbb{Z}_2$. For $|G| = 6$ we have $G \cong \mathbb{Z}_6, S_3$.

> *In many areas of mathematics, there are ways of 'building things up' and 'breaking things down'.*
>
> – Norman Block[38]

[37] This gets back to Grothendieck's contrast between the philosophy of chisel/hammer v. 'the rising sea'. See the endnotes of the Chapter 6 Notes. Put briefly: develop the encompassing algebraic theory/tools until your problem becomes trivially soluble.

[38] Attributed *N. Block. Abstract Algebra with Applications.* in Gallian's Contemporary Abstract Algebra.

9. Normal Subgroups

Let $K \leqslant G$. Our goal of this section is to make a group out of the cosets $\{Ka\}$. This means we need to define a group operation on the set of cosets.

Recall how we set up congruence in the integers as $a \mod n + b \mod n = (a+b) \mod n$. The operation being 'well-defined' depended upon a key result:

$$a \equiv b \mod n, c \equiv d \mod n \implies a + c \mod n = b + d \mod n$$

The subgroup version of this would be

$$a \equiv b \mod K, c \equiv d \mod K \implies ac \equiv bd \mod K$$

Unfortunately, this is not true in general. We'll only be concerned with subgroups that do satisfy such a property.

Let $K \leqslant G$. Recall that $Ka = \{ka : k \in K\}$ was called the *right coset*. Similarly, there is a left coset, $aK = \{ak : k \in K\}$. In general, $aK \neq Ka$.

Here's an example. Consider D_4 - the set of symmetries of the square. Let s be a reflection (say, over x-axis). Consider $K = \{e, s\} = \langle s \rangle \leqslant D_4$. Let t be reflection over $x = y$. Then we say that $t = r \circ s$, where r was the (elementary) rotation by 90°. Then $tK \neq Kt$, as

$$tK = \{te, ts\} = \{t, (rs)s\} = \{t, r\} \qquad Kt = \{et, st\} = \{t, s(rs)\} = \{t, r^3\}$$

Definition 9.1. A subgroup $N \leqslant G$ is *normal* if $Na = aN$ for every $a \in G$. Denoted $N \trianglelefteq G$.[39]

An observation: if $N \leqslant G$ and G is abelian then $an = na$ for every $a \in G$ and $n \in N$. Thus $Na = aN$. In particular, *every subgroup of an abelian group is normal*.

For any group G the center $Z(G) \leqslant G$ is always a normal subgroup, i.e. $Z(G) \trianglelefteq G$. This is because for $a \in G$ and $n \in Z(G)$ we have $an = na$ by definition.

Importantly, $aN = Na$ *does not imply that* $na = an$ *for every* $n \in N$.

Now the theorem:

Theorem 9.2. *Let* $N \trianglelefteq G$. *If* $a \equiv b \mod N$ *and* $c \equiv d \mod N$ *then* $ac \equiv bd \mod N$.

Proof. We want to show that $(ac)(bd)^{-1} \in N$. We have $ab^{-1} = m$ and $cd^{-1} = n$ for $m, n \in N$. We compute

$$\begin{aligned}(ac)(bd)^{-1} = (ac)(d^{-1}b^{-1}) &= a(nb^{-1}) && \text{by normality there exists } n' \in N \text{ so } b^{-1}n' = nb^{-1} \\ &= ab^{-1}n' \\ &= mn' \in N\end{aligned}$$

□

It is often useful to have alternative characterizations/methods-of-thinking about normality.

Theorem 9.3. *The following conditions on* $N \leqslant G$ *are equivalent:*

(1) $N \trianglelefteq G$
(2) $a^{-1}Na = N$ *for all* $a \in G$ *where* $a^{-1}Na = \{a^{-1}na : n \in N\}$

[39] Similar to subgroup notation $N \leqslant G$, you won't find the notation $N \trianglelefteq G$ in BH. However, it is featured in most algebra texts.

Proof. Let $N \trianglelefteq G$. Fix $a \in G$. Then $aN = Na$. We wish to show $a^{-1}Na = N$. Consider $a^{-1}na$. We have $na = an'$ by normality. Thus $a^{-1}(na) = a^{-1}(an') = n' \in N$. Now let $n \in N$. Then $n = (a^{-1}a)n(a^{-1}a) = a^{-1}(ana^{-1})a = a^{-1}n'a \in a^{-1}Na$. The last equality, that $ana^{-1} = n'$ follows from what we just proved. Thus $a^{-1}Na = N$.

Now suppose $a^{-1}Na = N$ for all $a \in G$. We want to show $N \trianglelefteq G$. Fix $b \in G$, we wish to show $bN = Nb$. For $bn \in bN$ we have $bnb^{-1} = n' \in N$ (apply hypothesis with $a = b^{-1}$). Thus $bn = n'b \in Nb$. For $nb \in Nb$ we have $b^{-1}nb = n' \in N$. Thus $nb = bn' \in bN$. So $bN = Nb$. Thus $N \trianglelefteq G$. □

Quotient Groups

Let $N \trianglelefteq G$. Let G/N denote the set of right cosets of N in G, i.e. $G/N = \{Na : a \in G\}$.

Our goal in this section is to define an operation on G/N and turn it into a group. All of our previous work in $\mathbb{Z}, F[x]$ and general rings suggests an approach:

$$(Na)(Nb) = Nab$$

We must verify that this is will defined (this is, of course, where normality will play a role and the work we did in the last section will come into play).

Theorem 9.4. *Let* $N \trianglelefteq G$. *If* $Na = Nc$ *and* $Nb = Nd$ *in* G/N *then* $Nab = Ncd$.

Proof. $Na = Nc$ implies $a \equiv c \mod N$. $Nb = Nd$ implies $b \equiv d \mod N$. Thus $ab \equiv cd \mod N$. Thus $Nab = Ncd$. □

The group G/N is called the *quotient group* of G by N. Pronounced "G mod N".

Theorem 9.5. *Let* $N \trianglelefteq G$. *Then*

(1) G/N *is a group with operation* $(Na)(Nb) = N(ab)$
(2) *if* G *is finite, then* $|G/N| = |G|/|N|$
(3) *If* G *is abelian then* G/N *is abelian*

Proof. (1) The operation is well-defined by Theorem 9.4. It is straightforward that it is closed. It is associative since the operation in G is associatve, i.e.: $Na(NbNc) = Na(Nbc) = Nabc = (Nab)Nc = (NaNb)Nc$. There are four group axioms to show. It is easy to verify that Ne is the identity and Na^{-1} is the inverse of Na.

(2) By Lagrange's Theorem: $|G| = [G : N]|N|$. There are $[G : N]$ distinct cosets, the number of distinct cosets. Thus $|G|/|N| = [G : N]$.

(3) $NaNb = Nab = Nba = NbNa$ for all $a, b \in G$.

□

There are quite a few examples provided in pages 256-259 in BH (D_4, $\mathbb{Z}_n$, etc) it is very useful to look at these and internalize them.

9.1. **The Structure of Groups.** As we said in Chapter 1, algebra is a study in structure. For $N \trianglelefteq G$ the structure of G,N and G/N are related[40], and if we want to understand the structure of any one in particular of these groups, we can extract information by investigating the other two. Here are some (simple) examples of this general principle.

Theorem 9.6. *Let* $N \trianglelefteq G$. *Then* G/N *is abelian if and only if* $aba^{-1}b^{-1} \in N$ *for all* $a, b \in G$.

[40] In particular they are related by *morphisms*, as we will see in 8.4! (and as we saw in rings)

Proof. G/N is abelian if and only if $NaNb = NbNa$ for all $a, b \in G$. By definition, this holds if and only if $Nab = Nba$. $Nab = Nba$ holds if and only if $ab(ba)^{-1} \in N$, i.e. if and only if $aba^{-1}b^{-1} \in N$. □

We have that $Z(G) \trianglelefteq G$.

Theorem 9.7. *If $G/Z(G)$ is cyclic, then G is abelian.*

Proof. Let $a, b \in G$. If $G/Z(G)$ is cyclic then $G/Z(G) = \langle Z(G)c \rangle$ for some c. Thus $Z(G)a = (Z(G)c)^k = Z(G)c^k$ and $Z(G)b = (Z(G)c)^m = Z(G)c^m$. Thus $a = zc^k$ and $b = z'c^m$ with $z, z' \in Z(G)$. Thus $ab = zc^k z'c^m = zz'c^{k+m} = z'c^k zc^m = ba$ □

Theorem 9.7 can be a bit of a strange to wrap your head around, as if $G/Z(G)$ is cyclic then G is abelian. But then $Z(G) = G$. Thus $G/Z(G) = \langle Ne \rangle$. However, it turns out to be very useful in small order group classification. You can also think about it as the contrapositive: if G is not abelian then $G/Z(G)$ is not cyclic.

The study of quotient groups will turn out to be equivalent to the study of homomorphisms of G - this will be similar to ring theory, as all normal subgroups will arise as kernels of some morphism, and we'll have a First Isomorphism Theorem (for groups).

10. Quotient Groups (via Fibers)

If $f : G \to H$ is a group homomorphism, recall that the fibers of f are the sets $f^{-1}(h)$ for $h \in H$. Notice that if f is surjective, then the fibers $\{f^{-1}(h)\}_{h \in H}$ partition G. See Figure 12 for a schematic of fibers for a surjective homomorphism $f : G \to H$.[41] Dummit and Foote (D+F) at [2] offers an excellent exposition in Chapter 3.

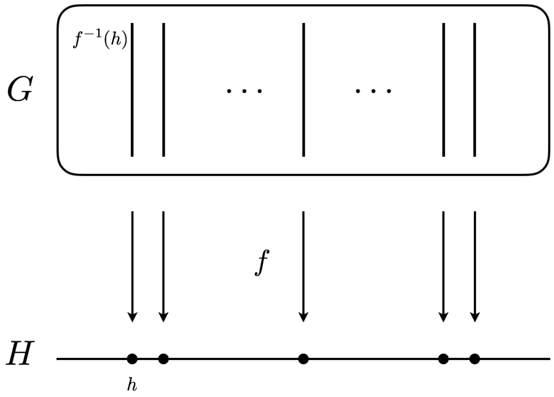


Figure 12. The fibers of f are the sets of elements of G which project to single elements of H, i.e. $f^{-1}(h)$ is the set of all elements b such that $b \mapsto h$. This is a schematic, where the vertical line in the box G above the point h depicts the fiber $f^{-1}(h)$.

[41] This image is used in Ch. 3 of Dummit and Foote. This picture shows the namesake of the term fiber. (This is also the quintessential schematic for *fiber bundle*; see [5] for a beautiful introduction/exposition in a physics setting).

The group operation in H provides a way to multiply two elements in the image of f - i.e. two elements on the horizontal. This suggests a natural product on the fibers lying above the two points - this would make the set of fibers into a group. See Figure 13. For $a, b \in H$ let $F_a = f^{-1}(a)$ (the fiber above a) and $F_b = f^{-1}(b)$ (the fiber above b). We'll define the product $F_a \cdot F_b = F_{ab}$, i.e. the fiber lying above $ab \in H$. This is associative since the operation in H is associative. The identity, will be F_{e_H} and the inverse of F_a will be $F_{a^{-1}}$.

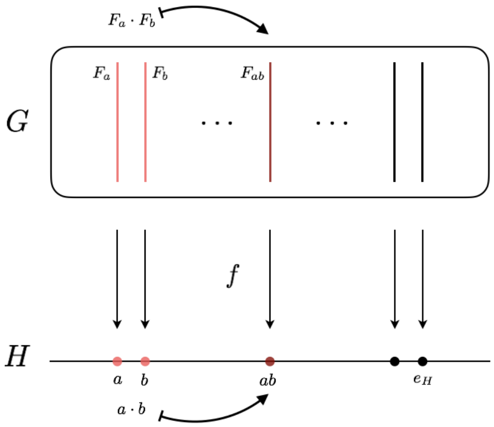


FIGURE 13. We make a group out of the fibers of f. For $a, b \in H$. We have a product $a \cdot b$, abbreviated ab. This suggests a product on fibers: $F_aF_b = F_{ab}$. We can then check the group axioms!

Let's do an example using this picture. Take $G = \mathbb{Z}$ and $H = \langle a \rangle$ with $|H| = n$, i.e. H is the cyclic group of order n. There is a natural homomorphism $f : G \to H$ given by

$$\mathbb{Z} \ni m \mapsto a^m$$

It is straightforward that f is a surjective homomorphism. Let's compute the fiber over a^k.

$$f^{-1}(a^k) = \{m \in \mathbb{Z} : a^m = a^k\} = \{m \in \mathbb{Z} : a^{m-k} = e_H\} = \{m \in \mathbb{Z} : m \equiv k \mod n\} = [k]_n$$

The second to last equality follows from Theorem 7.9, and fact that a has order n.

Thus F_{a^k}, the fiber of a^k, is the congruence class $[k]_n$. Now F_{a^i}, F_{a^j} be fibers over a_i and a_j respectively. We defined the operations on fibers as $F_{a^i} \cdot F_{a^j} = F_{a^i a^j} = F_{a^{i+j}}$. Since we know that the fibers are precisely congruence classes, this is equivalent to saying that $[i]_n \cdot [j]_n = [i + j]_n$. This is exactly how we set up modular arithmetic in $\mathbb{Z}_n$. See Figure 14.

What we've shown so far is that one can define a quotient group in terms of fibers. What we did previously was define a quotient group in terms of cosets. What's the relationship? Well the quotient group defined with fibers requires the homomorphism f explicitly, since the operation on fibers was defined by first projecting the fibers to H, using the product in H, then determining the fiber over this product. It is possible

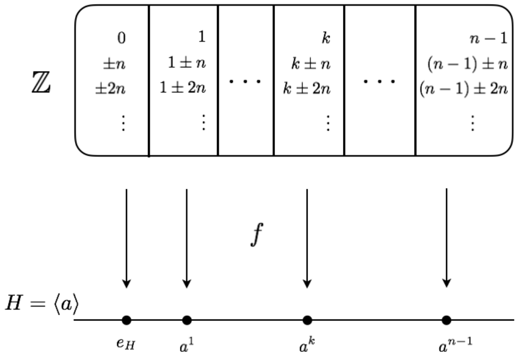


FIGURE 14. Here $G = \mathbb{Z}$ and $H = \langle a \rangle$ with order n, i.e. the cyclic subgroup of order n. The map is $m \mapsto a^m$. The fibers of a^k is precisely the congruence classes $[k]_n$. Thus the elements of the quotient group are the congruence classes, so this group is $\mathbb{Z}_n$! Furthermore, notice that the fibers are all just translates by $\langle n \rangle$, i.e. these are the cosets of the subgroup $\langle n \rangle$.

to define an operation on the fibers directly in terms of cosets, thus f will not enter explicitly. A natural question is then, are the fibers of a homomorphism the cosets of some subgroup?[42] Indeed, the fibers are the cosets of $\ker f$, i.e. a, b belong to the same fiber $f^{-1}h$ if and only if we have the coset equality $Ka = Kb$ where $K = \ker f$. Notice that another way of saying that a and b belong to the same fiber is to say $f(a) = f(b)$.

Theorem 10.1 (Lemma 8.19, [4]). *Let* $f : G \to H$ *be a group homomorphism with kernel* K. *Let* $a, b \in G$. *Then*

$$f(a) = f(b) \iff Ka = Kb$$

Proof. We have that

$$f(a) = f(b) \iff f(a)f(b)^{-1} = e_H \iff f(ab^{-1}) = e_H \iff ab^{-1} \in \ker f = K \iff Ka = Kb$$

□

That's slick. Let's spell this out a little more. Notice that if $ab^{-1} \in \ker f$ then $a = kb$ with $k \in \ker f$. Then $f(a) = f(kb) = f(k)f(b) = e_H f(b)$ so $f(a) = f(b)$. In other words, the product of b with any element of $\ker f$ has the same functional value $f(b) = f(a)$. So another way of thinking about $\ker f$ is as a container for all of the elements that when multiplied by an element don't change its functional value. The First Isomorphism Theorem for groups basically falls out of this perspective: it states we can then take G and mod out by $\ker f$ to get $\text{im}(f)$. But of course this happens, since by construction our quotient group was the collection of fibers $\{F_a\}$ which is in bijection correspondence with $\text{im}(f)$. Furthermore we saw that it turned out that if $b \in F_a$ then $F_a = Kb$ since the fiber of F_a consists of all elements that map to a, but thats precisely kb for each $k \in \ker f$.

Time for an example. Let $G = \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ (product group operation i.e. $(x, y) + (x', y') = (x + x', y + y')$. Let $H = \mathbb{R}$. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f((x, y)) = x$, i.e. projection on the x-axis (first coordinate). Then f is a homomorphism as $f\big((a, b) + (c, d)\big) = f\big((a +$

[42] This is precisely the content of WS8 prob 4.

$c, b+d)\big) = a + c = f(a,b) + f(c,d)$. A quick computation shows that

$$\ker f = \{(x,y) \in \mathbb{R}^2 : f(x,y) = 0\} = \{(0,y) : y \in \mathbb{R}\} = \text{ the } y\text{-axis}$$

Note $\ker f$ is a subgroup of $\mathbb{R}^2$ and the fiber of f over $a \in \mathbb{R}$ is the translate of the y-axis by the line $x = a$. This is precisely the coset $a + K$. See Figure 15.

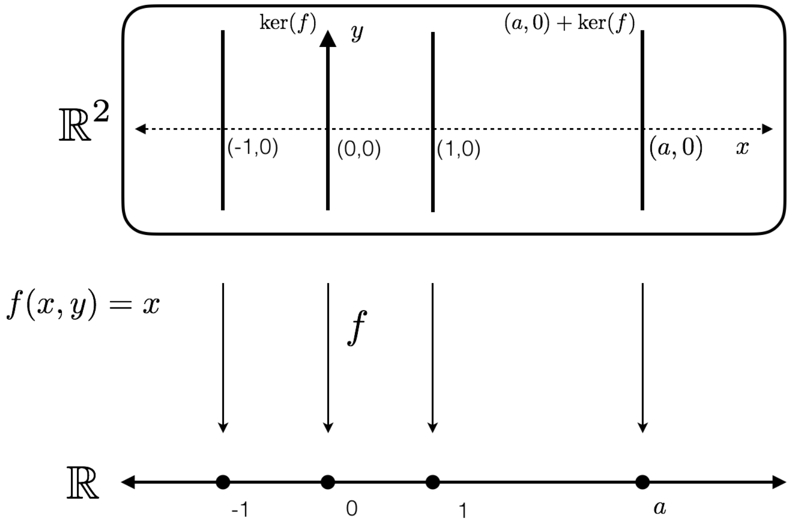


FIGURE 15. f is the projection of $\mathbb{R} \times \mathbb{R}$ onto the first coordinate $(x,y) \mapsto x$. $\ker f$ is the y-axis. The fibers are translates of $\ker f$, i.e. $a + \ker f$.

> *Mathematicians do not study objects, but relations among objects; they are indifferent to the replacement of objects by others as long as relations do not change. Matter is unimportant, only form interests them.*
>
> – Henri Poincaré[43]

8. Homomorphisms, Quotient Groups

We've seen the definition of kernel before. Recall:

Definition 8.1. Let $f : G \to H$ be a homomorphism of groups. Then the *kernel* of f is the set $\ker f = \{a \in G : f(a) = e_H\}$. Equivalently, $\ker f = f^{-1}(e_H)$, i.e. the fiber above e_H.

We saw that ideals corresponded to kernels in rings. We now have a similar result for groups.

Theorem 8.2. *Let* $f : G \to H$ *be a group homomorphism. Then* $\ker f \trianglelefteq G$.

Proof. We've seen that $K = \ker f$ is subgroup. We show that it is normal. By Theorem 8.11 it suffices to show $a^{-1}aK \subset K$. Let $a \in G$ and $c \in K$. We wish to show that $a^{-1}ca \in K$. We have $f(a^{-1}ca) = f(a^{-1})f(c)f(a) = f(a)^{-1}e_Hf(a) = f(a)^{-1}f(a) = e_H$. Thus $a^{-1}ca \in K$. So K is normal. □

[43] May be found in Gallian's Contemporary Abstract Algebra.

Theorem 8.2 says that every kernel is a normal subgroup.

Figure 16 is an image similar to one in Artin's Algebra, Chapter 5. It again depicts a schematic of a group homomorphism, but is closer to what we've drawn previously.

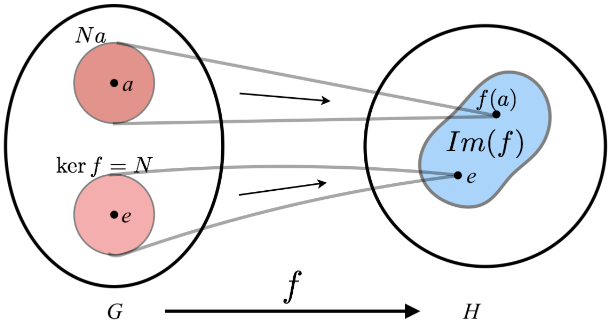


FIGURE 16. Schematic diagram of a homomorphism $f : G \to H$ with $\ker f = N$ which is the fiber over e and coset Na which is precisely the fiber over a.

Theorem 8.3. *Let* $f : G \to H$ *be a group homomorphism. Then* $\ker f = \langle e_G \rangle \iff f$ *is injective.*

Proof. Let $\ker f = \langle e_G \rangle$. Then $f(a) = f(b) \implies f(a)f(b)^{-1} = e_G \implies f(ab^{-1}) = e_G \implies ab^{-1} \in \ker f \implies a = b$. If f is injective then for $c \in \ker f$ we have $f(c) = e'_H = f(e_G) =$ implies $c = e_G$. □

The following theorem says that every normal subgroup is a kernel.

Theorem 8.4. *Let* $N \trianglelefteq G$. *Then the map* $\pi : G \to G/N$ *given by* $\pi(a) = Na$ *is a surjective homomorphism with* $\ker \pi = N$.

Proof. π is surjective since for any coset Na in G/N we have $\pi(a) = Na$. π is also a homomorphism as $\pi(ab) = Nab = NaNb = \pi(a)\pi(b)$. Then

$$\ker \pi = \{a \in G : \pi(a) = Ne\} = \{a \in G : Na = Ne\} = \{a \in G : ae^{-1} \in N\} = \{a : G : a \in N\} = N$$

□

π is sometimes called the *natural homomorphism*. Thus every normal subgroup arises as a kernel of the natural homomorphism π.

Lemma 8.5. *Let* $f : G \to H$ *be a group homomorphism with* $K = \ker f$. *Let* $a, b \in G$. *Then* $f(a) = f(b)$ *if and only if* $Ka = Kb$.

Proof. If $f(a) = f(b)$ then $f(a)f(b)^{-1} = e_H$ so $f(ab^{-1}) = e_H$. Thus $ab^{-1} \in \ker f = K$. Thus $Ka = Kb$.

If $Ka = Kb$ then $ab^{-1} \in K$ so $e_H = f(ab^{-1}) = f(a)f(b)^{-1}$. Thus $f(b) = f(a)$. □

Theorem 8.6 (First Isomorphism Theorem). *Let* $f : G \to H$ *be a surjective homomorphism. Then* $G/\ker f \cong H$.

Proof. Let $K = \ker f$. Define $\varphi : G/\ker f \to H$ by $\varphi(Ka) = f(a)$. We need to know φ is well-defined, i.e. its value depends only on the coset, not the representative. Suppose $Ka = Kb$. Then by Lemma 8.5 we have $f(a) = f(b)$. So $\varphi(Ka) = \varphi(Kb)$. So φ is well-defined.

To show φ is surjective, let $h \in H$. f is surjective so there is $a \in G$ with $f(a) = h$. Then $\varphi(Ka) = f(a) = h$. To see φ injective, suppose $\varphi(Ka) = \varphi(Kb)$. Then $f(a) = f(b)$. Thus $Ka = Kb$ by Lemma 8.5, so φ injective. Finally φ is a homomorphism as $\varphi(KaKb) = \varphi(Kab) = f(ab) = f(a)f(b) = \varphi(Ka)\varphi(Kb)$.

□

Here's an example. Let G, H be groups. Define $f : G \times H \to G$ by $f(a, b) = a$. Then f is a surjective homomorphism. $\ker f = \{(e_G, h) : h \in H\} = e_G \times H$. Thus $G \times H/e_G \times H \cong G$.

8.1. **Subgroups of Quotient Groups.** In this section we investigate the subgroups of the quotient group G/N. We do the Third Isomorphism Theorem for Groups, and some correspondences.

Theorem 8.7. *Let* $N \trianglelefteq G$. *Let* $N \leqslant K \leqslant G$. *Then* $K/N \leqslant G/N$.

Proof. We show $N \trianglelefteq K$. Since $Na = aN$ for every $a \in G$ we have $Na = aN$ for every $a \in K$. So $N \trianglelefteq K$ and K/N is a group. The elements of K/N are the cosets Na for $a \in K$. Thus $K/N \subset G/N$, and since K/N is a group itself we have $K/N \leqslant G/N$. □

Theorem 8.8 (Third Isomorphism Theorem)**.** *Let* $K, N \trianglelefteq G$ *with* $N \leqslant K \leqslant G$. *Then* $K/N \trianglelefteq G/N$ *and* $(G/N)/(K/N) \cong G/K$.

Proof. The idea is to define a surjective homomorphism from $G/N \to G/K$ with kernel K/N. The conclusion then follows by First isomorphism theorem.

Define $f : G/N \to G/K$ by $f(Na) = Ka$. We claim this is well-defined. If $Na = Nb$ then $ab^{-1} \in N \subset K$. Thus $Ka = Kb$. So f is well-defined.

f is surjective as for Ka we can choose Na and $f(Na) = Ka$. We compute $\ker f$.

$$\ker(f) = \{Na \in G/N : f(Na) = Ke\} = \{Na : Ka = K\} = \{Na : a \in K\} = K/N$$

Thus $K/N \trianglelefteq G/N$ since it arises as a kernel. Furthermore by First Iso. Thm $(G/N)/(K/N) = (G/N)/\ker f \cong G/K$

□

Theorem 8.9. *Let* $f : G \to H$ *be a group homomorphism. Let* $M \trianglelefteq H$ *and* $N = f^{-1}(M)$. *Then* $N \trianglelefteq G$.

Proof. Let $a \in G$ and $n \in N$. We show $a^{-1}na \in N$. We have $f(a^{-1}na) = f(a^{-1})f(n)f(a) \in M$ since $f(n) \in M$. Thus $N \trianglelefteq G$. □

The following two theorems are the 'correspondence theorems', which we saw previously for rings, and then in the Exam 2 prep.

Corollary 8.10 (Corollary 8.23 [4])**.** *Let* $N \trianglelefteq G$. *Let* $K \leqslant G$ *with* $N \subset K$. *Then* $K \trianglelefteq G$ *if and only if* $K/N \trianglelefteq G/N$.

Proof. Let $K \trianglelefteq G$. Then $K/N \trianglelefteq G/N$ by Third Iso. Thm. Now let $K/N \trianglelefteq G/N$. Then from the correspondence theorem $K = \pi^{-1}(K/N)$ where π is the natural homomorphism $\pi : G \to G/N$. $K = \pi^{-1}(K/N) \trianglelefteq G$. by Theorem 8.9.

□

We now have complete information about the subgroups of G/N that arise from subgroups of G that contain N.

Corollary 8.11 (Theorem 8.24 [4])**.** *If* T *is any subgroup of* G/N *then* $T = H/N$ *where* H *is a subgroup of* G *that contains* N*.*

The notion of a 'group', viewed only 30 years ago as the epitome of sophistication, is today one of the mathematical concepts most widely used in physics, chemistry, biochemistry, and mathematics itself.

– Alexei Sossinski[44]

9. Topics in Group Theory

Chapter 9 we look at some deep results in finite group theory. We address the classification of all finite abelian groups. We begin a path toward Sylow's Theorems, looking at the idea of conjugacies. We cover 9.1-9.2 and the beginning of 9.4.

10. Direct Products

The universe is an enormous direct product of representations of symmetry groups.

– Steven Weinberg[45]

Let G, H be groups. Then $G \times H$ is also a group as we've seen. In this section we look at the conditions under which a group is isomorphic to a direct product of certain subgroups.

Let $G_1, G_2, \ldots G_n$ be groups. Define coordinate-wise operation as

$$(a_1, a_2, \ldots, a_n)(b_1, b_2, \ldots, b_n) = (a_1b_1, a_2b_2, \ldots, a_nb_n)$$

It's easy to verify that $G_1 \times G_2 \times \ldots \times G_n$ is a group. Notice $(e_1, e_2, \ldots, e_n)$ is the identity and with inverses $(a_1^{-1}, a_2^{-1}, \ldots, a_n^{-1})$.

This group is called the *direct product* of $G_1, G_2, \ldots, G_n$. A remark on notation: when each G_i is abelian, that the direct product is often called the *direct sum* and denoted $G_1 \oplus G_2 \oplus \ldots G_n$.

Consider the group $\mathbb{Z}_6$. Notice that $M = \{0, 3\}$ and $N = \{0, 2, 4\}$ are subgroups. Every element of $\mathbb{Z}_6$ can be written as a sum of elements in M and N, i.e. $1 = 4 + 3$ and $5 = 3 + 2$. So $\mathbb{Z}_6 \cong M \times N$.

Lemma 10.1. *Let $M, N \trianglelefteq G$ such that $M \cap N = \langle e \rangle$. If $a \in M$ and $b \in N$ then $ab = ba$.*

Proof. Consider $a^{-1}b^{-1}ab$. M is normal so $b^{-1}ab \in M$. By closure of M we have $a^{-1}b^{-1}ab \in M$. We have $a^{-1}b^{-1}a \in N$. Closure implies $a^{-1}b^{-1}ab \in N$. Thus $a^{-1}b^{-1}ab \in M \cap N = \langle e \rangle$. Therefore $ab = ba$. □

Theorem 10.2. *Let $N_1, N_2 \ldots, N_k$ be normal subgroups of G such that every element in G can be written uniquely as $a_1a_2 \cdots a_k$ with $a_i \in N_i$. Then $G \cong N_1 \times N_2 \times \ldots \times N_k$.*

Remark 10.3. *Uniqueness here means if $a_1a_2 \cdots a_k = b_1b_2 \cdots b_k$ with $a_i, b_i \in N_i$ then $a_i = b_i$.*

Proof. Define a map

$$f : N_1 \times N_2 \times \cdots \times N_k \to G$$

by $f(a_1, a_2, \ldots, a_k) = a_1a_2 \cdots a_k)$ Every element of G can be written in that form, so f is surjective. If $a_1a_2 \cdots a_k = f(a_1, a_2, \ldots, a_k) = f(b_1, b_2, \ldots, b_k) = b_1b_2 \cdots b_k$

[44] Found in Gallian.

[45] Found in Gallian.

then $a_1a_2\cdots a_k = b_1b_2\cdots b_k$. By uniqueness this means $a_i = b_i$ for each i. Thus $(a_1, a_2, \dots, a_k) = (b_1, b_2, \dots, b_k)$. So f is injective.

It remains to show that f is a homomorphism. We want to apply Lemma 10.1. Notice that if $a \in N_i \cap N_j$ then a can be written as $a = e_1e_2\cdots e_{i-1}ae_{i+1}\cdots e_k = e_1e_2\cdots e_{j-1}ae_{j+1}\cdots e_k$. Here e_i just means e in the ith position. By uniqueness this forces $a = e$.

$$f\big((a_1,\dots,a_k)(b_1,\dots,b_k)\big) = f(a_1b_1,\dots,a_kb_k) = a_1b_1a_2b_2\cdots a_kb_k = a_1a_2\cdots a_kb_1b_2\cdots b_k$$

(This is done by commuting each a_i left)

Thus f is an isomorphism.

□

Depending upon the context we can think of G as the *external* direct product of the N_i, i.e. tuples $(a_1, a_2, \dots, a_k)$ or as an *internal* direct product where elements are written $a_1a_2\cdots a_k$.

If $M, N \leqslant G$ let $MN = \{mn : m \in N, n \in N\}$. The following theorem is often easier to apply than Theorem 10.2.

Theorem 10.4. *If* $M, N \trianglelefteq G$ *with* $G = MN$ *and* $M \cap N = \langle e \rangle$ *then* $G = M \times N$.

Proof. We have every element of G is of the form mn. Suppose $mn = m'n'$. Then

$$mn = m_1n_1$$
$$m_1^{-1}m = n_1n^{-1}$$

But $m_1^{-1}m \in M$ and $n_1n^{-1} \in N$. Thus $m_1^{-1}m = e = n_1n^{-1}$ so $m_1 = m$ and $n_1 = n$. This implies every element in G can be written uniquely in the form mn for $m \in M$ and $n \in N$. Thus $G = M \times N$ by Theroem 10.2.

□

11. Finite Abelian Groups

In this section we will classify *all* finite abelian groups! We shall prove that every finite abelian group G is a direct sum of cyclic subgroups and that their orders are determined by G.

All groups in this section will therefore be abelian. Thus we will use additive notation. BH has translations, and we'll review it here as well.

Multiplicative Notation	Additive Notation
ab	$a+b$
e	0
a^k	ka
MN	$M+N$
direct product $M \times N$	direct sum $M \oplus N$
direct factor M	direct summand M

We recall three theorems, now written in additive notation.

Theorem 11.1. *Let* G *be abelian. Let* $a \in G$*. Then*

(1) *If* $|a| = n$ *then* $ka = 0$ *if and only if* $n|k$.
(2) *If* a *has order* td *with* $d > 0$ *then* ta *has order* d.

Theorem 11.2. *If $N_1, \dots, N_k$ are normal subgroups of abelian group G such that each element of G can be written uniquely in the form $a_1 + a_2 \dots + a_k$ with $a_i \in N_i$. Then $G \cong N_1 \oplus N_2 \oplus \dots \oplus N_k$.*

Theorem 11.3. *If $M, N \leqslant G$ with $G = M + N$ and $M \cap N = \langle 0 \rangle$ then $G = M \oplus N$.*

Let G be abelian group and let p be prime. Let $G(p) = \{a \in G : |a| = p^n \text{ some } n \geqslant 0\}$.

Proposition 11.4. $G(p) \leqslant G$

Proof. $G(p)$ is nonempty as e has order $1 = p^0$. Let $a, b \in G(p)$ with orders p^k, p^m. Then $(p^k p^m)(a+b) = (p^m p^k)a + (p^k p^m)b) = 0$. $p^k(-a) = -p^k a = 0$. □

Here's example. Let $G = \mathbb{Z}_{12}$. Then $G(2)$ has elements of order $2^0, 2^1, \dots$. $G(2) = \{0, 3, 6, 0\}$.

Lemma 11.5. *Let G be abelian group. Let $a \in G$ with $|a| < \infty$. Then $a = a_1 + a_2 + \dots + a_t$ with $a_i \in G(p_i)$ where $p_1, \dots, p_t$ are distinct positive primes that divide the order of a.*

Proof. By Fundamental Theorem of Arithmetic $|a| = p_1^{n_1} p_2^{n_2} \cdots p_t^{n_t}$. We induct on the number of distinct primes that divide $|a|$, i.e. k.

The base case: if $|a| = p_1$, i.e. divisible by only a single prime, then the order of $|a|$ is prime, so $a \in G(p_1)$. Thus the Lemma holds.

Now assume the inductive hypothesis that the lemma is true for all elements whose order is divisible by at most $k-1$ distinct primes and that $|a|$ is divisible by the distinct primes $|a| = p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$ with each $n_i > 0$. Let $m = p_1^{n_1}$ and $n = p_2^{n_2} \cdots p_t^{n_t}$. Then $|a| = mn$. Then $(m, n) = 1$ and by Bezout's theorem there exist u, v with $1 = um + vn$. Thus $a = 1a = (mu + vn)a = mua + nva$.

But $mua \in G(p_1)$ since $p_1^{n_1}(mua) = nmua = u(nma) = 0$. Similarly $m(nva) = 0$. Thus $|nva| | m$. m has only $k-1$ distinct prime divisors. Thus by the inductive hypothesis $nva = a_2 + a_3 + \dots + a_k$. with $a_i \in G(p_i)$. Let $a_1 = mua$ then $a = mua + nva = a_1 + a_2 + \dots + a_k$ with $a_i \in G(p_i)$.

□

Theorem 11.6. *If G is a finite abelian group, then*

$$G = G(p_1) \bigoplus G(p_2) \bigoplus \cdots \bigoplus G(p_t)$$

where $p_1, \dots, p_t$ are distinct positive primes that divide the order of G.

Proof. If $a \in G$, then $|a| | |G|$ [46]. Hence $a = a_1 + a_2 + \dots + a_t$ with $a_i \in G(p_i)$ by Lemma 11.5. Here $a_j = 0$ if $p_j \not| |a|$.

We prove this expression is unique, then apply Theorem.

If $a_1 + a_2 + \dots + a_t = b_1 + b_2 + \dots b_t$ with $a_i, b_i \in G(p_i$ then $a_1 - b_1 = (b_2 - a_2) + (b_3 - a_3) + \dots + (b_t - a_t)$ and $b_i - a_i \in G(p_i)$ thus has order $p_i^{r_i}$ (i.e. some power of p_i). If $m = p_2^{r_2} \cdots p_t^{r_t}$ then $m(b_i - a_i) = 0$ for $i \geqslant 2$, so that

$$m(a_1 - b_1) = m(b_2 - a_2) + m(b_3 - a_3) + \dots + m(b_t - a_t) = 0$$

Thus the order of $|a_1 - b_1| | m$.

But $a_1 - b_1 \in G(p_1)$ so its order is a power of p_1. The only power of p_1 that divides $m = p_2^{r_2} \cdots p_t^{r_t}$ is $p_1^0 = 1$. Thus $a_1 - b_1 = 0$ and $a_1 = b_1$. Similar arguments apply for any

[46] This was a corollary to Lagrange.

i. Thus every element can be written in the form $a_1 + \ldots a_t$ with $a_i \in G(p_i)$ and thus $G = G(p_1) \oplus \cdots \oplus G(p_t)$ by Theorem 10.2.

□

If p is prime then a group in which every element has order p^n some n is called a p-group. Thus each of the $G(p_i)$ is a p-group. An element a of p-group B is called *an element of maximal order* if $|b| \leqslant |a|$ for all $b \in B$.

If $|a| = p^n$ and $b \in B$ with $|b| = p^j$ then $j \leqslant n$. $p^n = p^j p^{n-j}$ thus $p^n b = p^{n-j} p^j b = 0$. Thus if a is an element of maximal order p^n in a p-group B then $p^n b = 0$ for every $b \in B$.

Elements of maximal order always exist in a finite p-group.

Lemma 11.7 (Lemma 9.6 [4]). *Let G be a finite abelian p-group and a an element of maximal order in G. Then there is a subgroup $K \leqslant G$ such that $G = \langle a \rangle \bigoplus K$.*

Proof. Proof in BH. □

Theorem 11.8. *Every finite abelian group G is the direct sum of cyclic groups, each of prime power[47] order.*

Proof. $G = G(p_1) \oplus G(p_2) \oplus \ldots \oplus G(p_t)$ one for each $p_i||G|$ by Theorem 11.6. Each $G(p)$ is a p-group.

So it remains to show that every finite abelian p-group H is a direct sum of cyclic groups, each of order a power of p.

We prove this by induction on the order of H. The base case is when $|H| = 2$. Then H is cyclic.[48]

Now assume inductively (i.e. let the inductive hypothesis be) that it is true for p-groups whose order is less than $|H|$.[49] Let a be an element of maximal order p^n in H. The $H = \langle a \rangle \bigoplus K$. Then $H = \langle a \rangle \oplus K$ by Lemma 11.7. Notice that K is a p-group and $|K| < |H|$. Thus by induction, K is a direct sum of cyclic groups, each with order a power of p.

□

The number 36 can be written as a product of prime powers in four ways: $36 = 2 \times 2 \times 3 \times 3 = 2 \times 2 \times 3^2 = 2^2 \times 3 \times 3 = 2^2 \times 3^2$. Thus Theorem 11.8 every abelian group of order 36 must be isomorphic to one of the following groups:

$$\mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_3 \qquad \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_9 \qquad \mathbb{Z}_4 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_4 \qquad \mathbb{Z}_4 \oplus \mathbb{Z}_9$$

One can verify that no two of the groups are isomorphic (for instance, check the numbers of elements of order 2 or 3).

Lemma 11.9. *If $(m, k) = 1$ then $\mathbb{Z}_m \bigoplus \mathbb{Z}_k \cong \mathbb{Z}_{mk}$.*

Proof. The order of $(1, 1)$ in $\mathbb{Z}_m \oplus \mathbb{Z}_k$ is the smallest positive integer t such that $(0, 0) = t(1, 1) = (t, t)$. Then $t \equiv 0 \mod m$ and $t \equiv 0 \mod k$ so $k|t$ and $m|t$. But $(m, k) = 1$ implies $mk|t$ (least common multiple), so $mk \leqslant t$. But t is the smallest such integer, thus $mk = t = |(1, 1)|$. Therefore $\mathbb{Z}_m \oplus \mathbb{Z}_k$ is a cyclic group of ordered nm generated by $(1, 1)$ and isomorphic to $\mathbb{Z}_{mk}$. □

[47] Meaning the order of a p^k with p prime

[48] All groups of prime order are cyclic by Theorem 8.7 in BH

[49] i.e. The inductive hypothesis is that every p-group of order less than $|H|$ is a direct sum of cyclic groups, each of order a power of the prime p.

Theorem 11.10. *If* $n = p_1^{n_1} p_2^{n_2} \cdots p_t^{n_t}$ *with* $p_1, \dots p_t$ *distinct primes, then*

$$\mathbb{Z}_n \cong \mathbb{Z}_{p_1^{n_1}} \bigoplus \mathbb{Z}_{p_2^{n_2}} \bigoplus \cdots \bigoplus \mathbb{Z}_{p_t^{n_t}}$$

Proof. We'll do induction on the order of the group. The base case when $n = 2$ is true.

Assume that it is true for groups of order less than n. Apply Lemma 11.9 with $m = p_1^{n_1}$ and $k = p_2^{n_2} \cdots p_t^{n_t}$. Then $\mathbb{Z}_n \cong \mathbb{Z}_{p_1^{n_1}} \oplus Z_k$. By the inductive hypothesis $\mathbb{Z}_k \cong Z_{p_2^{n_2}} \oplus \cdots \oplus \mathbb{Z}_{p_t^{n_t}}$. □

We now combine Theorems 11.10 and 11.8. This implies second a way of expressing a finite abelian group as a direct sum of cyclic groups.

Consider the group

$$G = \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_4 \oplus \mathbb{Z}_8 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_3 \oplus \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_5 \oplus \mathbb{Z}_25$$

Arrange the prime power orders of the cyclic factors by size with one row for each prime:

2	2	$2^2 = 4$	$2^3 = 8$
	3	3	3
		5	$5^2 = 25$
2	6	60	600

Now rearrange the cyclic factors of G using the columns, and apply Theorem 11.10. Then

$$G \cong \mathbb{Z}_2 \oplus \mathbb{Z}_6 \oplus \mathbb{Z}_{60} \oplus \mathbb{Z}_{600}$$

The last decomposition is sometimes more convenient. There are fewer factors and the order of each cyclic factor divides the order of the next one. This leads to the next theorem:

Theorem 11.11. *Every finite abelian group is the direct sum of cyclic groups of orders* $m_1, m_2, \dots, m_t$ *where* $m_1 | m_2 | m_3 | \cdots | m_{t-1} | m_t$.

If G is finite abelian group then the integers $m_1, \dots, m_t$ in Theorem 11.11 are called the *invariant factors* of G. When G is written as direct sum of cyclic groups of prime power orders, the prime powers are called the *elementary divisors* of G.

Here's an example. All abelian groups of order 36 can be classified up to isomorphism in terms of their elementary divisors or their invariant factors:

Group	Elementary Divisors	Invariant Factors	Isomorphic Group
$\mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_3$	$2, 2, 3, 3$	$6, 6$	$\mathbb{Z}_6 \oplus \mathbb{Z}_6$
$\mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_9$	$2, 2, 3^2$	2,18	$\mathbb{Z}_2 \oplus \mathbb{Z}_{18}$
$\mathbb{Z}_4 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_3$	$2^2, 3, 3$	$3, 12$	$\mathbb{Z}_3 \oplus \mathbb{Z}_{12}$
$\mathbb{Z}_4 \oplus \mathbb{Z}_9$	$2^2, 2^3$	36	$\mathbb{Z}_{36}$

11.1. **The Upshot of the Fundamental Theorem.** From the text in Hungerford, it may be difficult to discern the upshot of this theorem. We review some of Gallian's exposition in *Contemporary Abstract Algebra* on the Fundamental Theorem. The Fundamental Theorem of Finite Abelian Groups describes (up to isomorphism) all finite abelian groups in a standard way. In words, the theorem says

> *Every finite abelian group is a direct product of cyclic groups of prime-power order. Moreover, the numbers of terms in the product and the orders of the cyclic groups are uniquely determined by the group.*

We have a classification of cyclic groups - we know a cyclic group of order n is isomorphic to $\mathbb{Z}_n$. Thus every finite abelian group G is isomorphic to a group of the form

$$\mathbb{Z}_{p_1^{n_1}} \oplus \mathbb{Z}_{p_2^{n_2}} \oplus \ldots \oplus \mathbb{Z}_{p_k^{n_k}}$$

where the p_i are not necessarily distinct primes and the prime powers $p_1^{n_1}, \ldots, p_k^{n_k}$ are uniquely determined by G. Writing a group in this form is *determining the isomorphism class of* G - that is, we've expressed an isomorphic group to G, in a standard way.

For now, let's look at abelian groups with order p^k for p prime. In general, there is one group of order p^k for each set of positive integers whose sum is k (this is called a partition), i.e. if k can be written as $k = n_1 + n_2 \ldots + n_t$ where each $n_i > 0$. Then $Z_{p^{n_1}} \oplus \mathbb{Z}_{p^{n_2}} \oplus \ldots \oplus \mathbb{Z}_{p^{n_t}}$ is a group of order p^k. So for $k = 3$ we have partitions $3, 2 + 1, 1 + 1 + 1$, for these three partitions we have groups

$$\mathbb{Z}_{p^3}, \mathbb{Z}_{p^2} \oplus \mathbb{Z}_p, \mathbb{Z}_p \oplus \mathbb{Z}_p \oplus \mathbb{Z}_p$$

How do we know these groups are not isomorphic? (Hint compare the order of elements of maximal order). The first has an element of order p^3. The second an element of order p^2 (and any element raised to p^2 is 0) and the third only elements of order p (and all elements raised to p are 0).

Now let's move on to arbitrary abelian groups, say of order n. When n has two or more distinct prime divisors then we write n in its prime-power decomposition[50], say $n = p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$. Then we form all abelian groups of order $p_1^{n_2}$ then $p_2^{n_2}$, and so on. Then we form all possible external products of these groups. For instance, let $n = 1176 = 2^3 \cdot 3 \cdot 7^2$. Then the complete list of distinct isomorphism classes of abelian groups of order 1176 is:

$$\begin{aligned}
\mathbb{Z}_8 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_{49} \\
\mathbb{Z}_4 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_{49} \\
\mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_{49} \\
\mathbb{Z}_8 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_7 \oplus \mathbb{Z}_7 \\
\mathbb{Z}_4 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_7 \oplus \mathbb{Z}_7 \\
\mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_7 \oplus \mathbb{Z}_7
\end{aligned}$$

12. Conjugacy

In this section we discuss conjugacy. Most of the exposition here is drawn from 9.4. Some is from Herstein [3] as Chapter 2, Section 11: Another Counting Principle[51]. In this section Herstein points out that conjugacy can be thought of as a counting principle. A favorite way of counting in mathematics is to count up a situation in two different ways, the comparison of the two counts then can be used as a means for drawing conclusions. Taking from Herstein:

[50] via Fundamental Theorem of Arithmetic

[51] Be careful if looking at Herstein - his notation in different than Hungerfords

> *Generally speaking, one introduces an equivalence relation on a finite set, measures the size of the equivalence classes under this relation, and then equates the number of elements in the set to the sum of the orders of these equivalence classes.*

Let G be a group. Let $a, b \in G$. We say that a is *conjugate* to b if there exists $x \in G$ such that $b = x^{-1}ax$.

We'll use the idea of equivalence class here, which can be found in Appendix D.

Theorem 12.1. *Conjugacy is an equivalence relation on* G.

Proof. We write $a \sim b$ if a is conjugate to b.

(1) Reflexive. $a \sim a$ since $a = e^{-1}ae$.
(2) Symmetric. Let $a \sim b$. Then $b = x^{-1}ax$. Thus $xbx^{-1} = a$ so $a = (x^{-1})^{-1}b(x^{-1})$.
(3) Transitive. Let $a \sim b$ and $b \sim c$. Then $b = x^{-1}ax$ and $c = y^{-1}by$ so $c = y^{-1}(x^{-1}ax)y = (xy)^{-1}axy$. So $a \sim c$.

□

The equivalence classes of G under the relation of conjugacy are called *conjugacy classes*. Thus the conjugacy class of $a \in G$ is $\{b \in G : a \sim b\}$. Standard results imply that two conjugacy classes are either disjoint or equal and that the group is the union of all its distinct conjugacy classes.

We'll now show that the size of the conjugacy class of a is the index of a certain subgroup. To get a feeling for this, notice that we can rewrite the conjugacy class as

$$\{b \in G : a \sim b\} = \{b \in G b = x^{-1}ax \text{ for some } x \in G\} = \{x^{-1}ax : x \in G\}$$

Alright, so the conjugacy class of a is just the set of conjugates of a by every element in the group. What happens for elements $g \in G$ such that $g^{-1}ag = a$?

Let G be a group. The *centralizer* of a is all elements of G that commute with a, and is denoted $C(a)$, i.e.

$$C(a) = \{g \in G : ga = ag\}$$

Proposition 12.2. *If* G *is a group and* $a \in G$ *then* $C(a) \leqslant G$.

Proof. (1) $C(a)$ is nonempty. $e \in C(a)$ since $ea = ae$.
(2) Closed. Let $x, y \in C(a)$. Then $(xy)a = xyaxay = axy = a(xy)$.
(3) Inverses. Let $x \in C(a)$. Then $xa = ax$. Multiply x^{-1} on right to get $xax^{-1} = a$ and then on left to get $ax^{-1} = x^{-1}a$.

□

Theorem 12.3. *Let* G *be a finite group. Let* $a \in G$. *The number of elements in conjugacy class of* a *is* $[G : C(a)]$ *and this number divides* $|G|$.

Proof. Let S be the set of distinct right cosets of $C(a)$.

Let T be the conjugacy class of a in G.

Define a function $f : S \to T$ by

$$f(C(a)x) = x^{-1}ax$$

We show that f is well-defined bijection (of sets). If so, then $|T| = |S| = [G : C(a)]$, which divides $|G|$ by Lagrange's Theorem.

We show with an $\iff$ proof that f is both well-defined and injective.

$$\begin{aligned} Cx = Cy &\iff xy^{-1}C \\ &\iff (xy^{-1})a = a(xy^{-1}) \text{ this follows from the definition of } C(a) \\ &\iff a = (xy^{-1})^{-1}a(xy^{-1}) \\ &\iff a = yx^{-1}axy^{-1} \\ &\iff y^{-1}ay = x^{-1}ax \\ &\iff f(Cy) = f(Cx) \end{aligned}$$

Reading the proof forward shows well-defined. Going backward shows injective.

Now for surjective - given any conjugate $u^{-1}au$ of a we have that $f(Cu) = u^{-1}au$. Thus f is bijective.

□

Now for some discussion. Let G be finite. Let $C_1, C_2, \ldots, C_t$ be the distinct conjugacy classes of G. Then

$$G = C_1 \cup C_2 \cup \ldots \cup C_t$$

Since distinct conjugacy classes are mutually disjoint we have

$$|G| = |C_1 \cup \ldots \cup C_t| = |C_1| + \ldots + |C_t|$$

Now choose $a_i \in C_i$. Then C_i consists of the conjugates of a_i. By the Theorem 12.3 we have $|C_i| = [G : C(a_i)]$. This turns the above equation into

$$|G| = [G : C(a_1)] + \ldots + [G : C(a_t)]$$

This equation is called the *class equation* of the group G. It is quite powerful.[52]

Notice that if $c \in Z(G)$, then $cx = xc$ for all $x \in G$. Thus $c = x^{-1}cx$ for all x, so that the conjugacy class of c consists of only one element, i.e. $\{c\}$. Therefore we can rewrite the class equation in the following form

$$|G| = |Z(G)| + |C_1| + \ldots |C_r|$$

where $C_1, \ldots, C_r$ are distinct conjugacy classes of G that contain more than one element and each $|C_i|$ divides $|G|$.

Corollary 12.4 (Herstein [3], 2.11.2). *If* $|G| = p^n$ *where* p *is prime then* $Z(G) \neq \langle e \rangle$.

Proof. The last version of the class equation gives

$$|G| = |Z(G)| + |C_1| + \ldots |C_r|$$

$|G| = p^n$ and $|C_i|$ divides p^n, thus $|C_i| = p^{n_i}$ for some $0 < n_i < n$. Therefore we have

$$p^n = |Z(G)| + p^{n_1} + p^{n_2} \ldots + p^{n_r}$$

Now p divides p^n and p divides $p^{n_1} + p^{n_2} \ldots + p^{n_r}$ since $n_i > 0$. Thus p divides $p^n - (p^{n_1} + p^{n_2} \ldots + p^{n_r}) = |G| - (|C_1| + \ldots |C_r|) = |Z(G)|$. Thus $Z(G) \neq \langle e \rangle$. □

[52] For instance, it is used to prove Sylow theorems. See [3] 2.11 for more applications.

Notes

[1] In Chapters 1-2 we worked with the rings $\mathbb{Z}$ and $\mathbb{Z}_n$; in Chapters 3 and 4 with the polynomial rings $F[x]$. In Chapter 6 (and some in Chapter 3) we have abstracted away all of the idiosyncrasies of these particular rings. We've built a very general and powerful framework with which to prove results about all rings (subsuming most results about the specific examples we've worked on). This is a widespread philosophy in mathematics, but runs particularly strongly through algebra. Perhaps some interesting insight can be found in Colin McLarty's essay on Alexander Grothendieck's philosophy of 'la mer qui monte' or 'the rising sea'.[53] I will quote him putting Grothendieck's remarks in context.

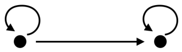
> *Grothendieck describes two styles in mathematics. If you think of a theorem to be proved as a nut to be opened, so as to reach "the nourishing flesh protected by the shell", then the hammer and chisel principle is: "put the cutting edge of the chisel against the shell and strike hard. If needed, begin again at many different points until the shell cracks - and you are satisfied". He says:*
>
> > *I can illustrate the second approach with the same image of a nut to be opened. The first analogy that came to my mind is of immersing the nut in some softening liquid, and why not simply water? From time to time you rub so the liquid penetrates better, and otherwise you let time pass. The shell becomes more flexible through weeks and monthswhen the time is ripe, hand pressure is enough, the shell opens like a perfectly ripened avocado!*
> >
> > *A different image came to me a few weeks ago. The unknown thing to be known appeared to me as some stretch of earth or hard marl, resisting penetration. . . the sea advances insensibly in silence, nothing seems to happen, nothing moves, the water is so far off you hardly hear it. . . yet it finally surrounds the resistant substance.*
>
> *In this "rising sea" the theorem is "submerged and dissolved by some more or less vast theory, going well beyond the results originally to be established"*

In many cases in Chapter 2 and 5, we've gone after problems with a 'hammer and chisel'. Chapter 6 manifests the 'rising sea' philosophy.

[2] Here's a small primer on category theory. We may expand this endnote into a section of notes in its own right. Category theory is a very primitive language for mathematics. That is one reason it is so powerful. Here are some familiar examples of categories. Rings together with ring homomorphisms form a category denoted $\mathbf{Ring}$. Commutative rings form a category $\mathbf{CRing}$ (a 'subcategory' of $\mathbf{Ring}$). Groups with group homomorphisms form a category $\mathbf{Grp}$. Topological spaces with continuous functions form a category $\mathbf{Top}$. Vector spaces over a field k together with linear maps $\mathbf{Vec}_k$.

Categories provide an organizing principle: specify a collection of objects and the appropriate 'structure-preserving' mappings, or *morphisms* between those objects. One reason this is elegant is that you specify only what is needed, i.e. the appropriate 'structure' you're interested in, and leave the rest behind. Nothing too abstract is going on here; only very disciplined organization.

A good way to visualize categories is as graphs with objects as the vertices and morphisms as the edges between vertices. Morphisms are subject to a composition relation, which specifies when following one path of edges is equivalent to following another path. Here we can visualize a small category:



The self edges here represent the identity morphism, (the identity map), which we'll see in the axioms for a category below. Here's a formal definition:

Definition 12.5. *A* category C *consists of a collection of objects, denoted* $\mathrm{obj}(C)$ *and a set of morphisms* $\mathrm{Hom}_C(a,b)$ *between any two objects* $a, b \in \mathrm{obj}(C)$*. An single morphism* $f : a \to b$ *is also called an* arrow *since it points/maps* $a \to b$*. The collection* C *must satisfy the following:*

(1) *Any two morphisms, say* $f \in \mathrm{Hom}_C(a,b)$ *and* $g \in \mathrm{Hom}_C(b,c)$*, can be composted to get another morphism* $g \circ f \in \mathrm{Hom}_C(a,c)$*.*

(2) *Composition is associative, i.e.* $h \in \mathrm{Hom}_C(c,d)$ *then* $h \circ (g \circ f) = (h \circ g) \circ f$

[53] C. McLarty. http://case.edu/artsci/phil/RisingSea.pdf

(3) *For each object* $\mathtt{b} \in \mathtt{obj}(\mathtt{C})$ *there is an identity morphism* $\mathtt{id}_\mathtt{b} \in \mathtt{Hom}_\mathtt{C}(\mathtt{b},\mathtt{b})$ *such that* $\mathtt{id}_\mathtt{b} \circ \mathtt{f} = \mathtt{f}$ *and* $\mathtt{g} \circ \mathtt{id}_\mathtt{b} = \mathtt{g}$ *for* $\mathtt{f} \in \mathtt{Hom}_\mathtt{C}(\mathtt{a},\mathtt{b})$ *and* $\mathtt{g} \in \mathtt{Hom}_\mathtt{C}(\mathtt{b},\mathtt{c})$

References

[1] Atiyah, M. (2002). *Mathematics in the 20th Century.* Bulletin of the London Mathematical Society, 34(1), 1-15.

[2] Dummit, D. S., & Foote, R. M. (2004). *Abstract algebra (Vol. 3).* Wiley.

[3] Herstein, I. (1975). Topics in Algebra. John Wiley & Sons.

[4] Hungerford, T. W. (2012). *Abstract Algebra: An Introduction.* Cengage Learning.

[5] Penrose, R. (2004). *The road to reality: A complete guide to the physical universe.* Jonathan Cape.

[6] Strogatz, S. (2010). *Group Think*. https://opinionator.blogs.nytimes.com/2010/05/02/group-think/.