Introduction to Ring Theory (MATH 228) – v 1.15

Paul Buckingham

About these notes

These notes provided the core material for an introductory course on ring theory once taught at the University of Alberta. The topics have since been moved to new courses.

Contents

(I) Fundamental Concepts

I – 1 Sets

Informally, a *set* is a collection of objects, called the *elements* of the set. There are three common ways to specify a set. Let us illustrate each by example:

(i) (a) Let X be the set of odd integers between 0 and 10.

(b) Let Y be the set of positive odd integers (an infinite set).

(ii) (a) Let $X = \{1, 3, 5, 7, 9\}$.

(b) Let $Y = \{1, 3, 5, 7, 9, \ldots\}$.

(iii) (a) Let $X = \{a \in \mathbb{Z} \mid a \text{ is odd and } 0 < a < 10\}$.

(b) Let $Y = \{a \in \mathbb{Z} \mid a \text{ is odd and } a > 0\}$.

We have used the symbol $\mathbb{Z}$ here. This symbol itself denotes a set, namely, the set of integers. That is,

$$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}.$$

The set of integers is one of the most important sets in the course, and arguably in all mathematics.

The symbol $\in$ in example (iii) means "is an element of", and the symbol $\mid$ may be read as "where". Thus, in example (iii)(a), the set X may be read as "the set of integers a where a is odd and $0 < a < 10$".

Another important set is the set $\mathbb{Q}$ of rational numbers, that is,

$$\mathbb{Q} = \{a/b \mid a, b \in \mathbb{Z} \text{ and } b \neq 0\}.$$

In words, we would say "the set of a/b where a and b are integers and $b \neq 0$".

It is also useful to be able to specify when something is *not* an element of a given set, in which case we use the symbol $\notin$, as in $\sqrt{2} \notin \mathbb{Q}$ ("$\sqrt{2}$ is not an element of $\mathbb{Q}$").

Cardinality

If X is a set, its *cardinality* $|X|$ is defined in this course to be

- the number of elements in X if X is finite (i.e., has finitely many elements),
- ∞ otherwise.

This is a simplified definition of cardinality, but we will not need anything more sophisticated in this course.

For example, $|\{3, \pi, \sqrt{2}, -e^3\}| = 4$, and $|\mathbb{Q}| = \infty$. Note that the set $\{1, 2, 2, 2, 6\}$ has cardinality 3, not 5, because its elements are 1, 2, and 6. This last set could also be written $\{1, 2, 6\}$, or $\{6, 2, 1\}$, $\{6, 1, 2\}$, etc.

Empty set

The *empty set* is the set with no elements. It is denoted either $\emptyset$ or $\{\}$.

Subsets

If X and Y are sets, then X is said to be a *subset* of Y if every element of X is an element of Y. If this is the case, then we write either $X \subseteq Y$ or $Y \supseteq X$. For example, $\{3, 7, -14\} \subseteq \mathbb{Z}$.

Operations on sets

Let Ω be a set and X, Y subsets of Ω. Then

- $X \cup Y = \{a \in \Omega \mid a \in X \text{ or } a \in Y\}$. (union of X and Y)
- $X \cap Y = \{a \in \Omega \mid a \in X \text{ and } a \in Y\}$. (intersection of X and Y)
- $X \smallsetminus Y = \{a \in X \mid a \notin Y\}$. (complement of Y in X)

For example, if $\Omega = \mathbb{Z}$, $X = \{3, 7, 9, 10\}$, and $Y = \{7, 9, 11, 15, 16\}$, then

$$\begin{aligned} X \cup Y &= \{3, 7, 9, 10, 11, 15, 16\} \\ X \cap Y &= \{7, 9\} \\ X \smallsetminus Y &= \{3, 10\} \end{aligned}$$

Note also that $\{1, 3\} \cap \{2, 5, 7\} = \emptyset$.

Cartesian product

If X and Y are sets, we can form the *cartesian product*

$$X \times Y = \{(x, y) \mid x \in X \text{ and } y \in Y\}.$$

For example, if $X = \{3, 7, 9, 10\}$ and $Y = \{3, 7, 11\}$, then

$$\begin{aligned} X \times Y = \{&(3,3), (3,7), (3,11), (7,3), (7,7), (7,11) \\ &(9,3), (9,7), (9,11), (10,3), (10,7), (10,11)\}. \end{aligned}$$

If X and Y are finite, then $|X \times Y| = |X|\,|Y|$.

We also define $X^2 = X \times X$, where X is any set. Iterating, we may define

$$X^n = \underbrace{X \times \cdots \times X}_{n}$$

for all $n \geq 1$.

I – 2 Functions

If X and Y are sets, a *function* (or *map*) is a rule f that assigns to each element x of X one, and only one, element of Y, denoted $f(x)$. The set X is called the *domain* of f, and Y is called the *codomain*. We write $f : X \to Y$ to specify that f is a function with domain X and codomain Y.

The symbol $\mapsto$ is used to specify where an element of the domain is mapped to under the given function. For example, we have the following functions:

$$\begin{aligned} f : \mathbb{Z} &\to \mathbb{Q} \\ a &\mapsto a/2 \quad \text{(i.e., } f(a) = a/2\text{)} \end{aligned}$$

$$\begin{aligned} g : \mathbb{R}_{>0} &\to \mathbb{R} \\ x &\mapsto \ln(x) \quad \text{(i.e., } g(x) = \ln(x)\text{, the natural logarithm of } x\text{)} \end{aligned}$$

$$\begin{aligned} h : \mathbb{Z}^2 &\to \mathbb{R} \\ (a, b) &\mapsto a + b\sqrt{2} \quad \text{(i.e., } h(a, b) = a + b\sqrt{2}\text{)} \end{aligned}$$

Image

If $f : X \to Y$ is a function, its image is

$$\text{Image}(f) = \{y \in Y \mid \text{there exists } x \in X \text{ such that } f(x) = y\} = \{f(x) \mid x \in X\}.$$

For example, if we define $f : \mathbb{R} \smallsetminus \{0\} \to \mathbb{R}$ by $x \mapsto x^2$,

$$\text{Image}(f) = \{f(x) \mid x \in \mathbb{R} \smallsetminus \{0\}\} = \{x^2 \mid x \in \mathbb{R} \smallsetminus \{0\}\} = \{y \in \mathbb{R} \mid y > 0\} = \mathbb{R}_{>0}.$$

Composing functions

If $f : X \to Y$ and $g : Y \to Z$ are functions, then their composition $g \circ f$ is the function

$$\begin{aligned} g \circ f : X &\to Z \\ x &\mapsto g(f(x)). \end{aligned}$$

For example, if

$$\begin{aligned} f : \mathbb{Z} \smallsetminus \{0\} &\to \mathbb{Z}_{>0} \\ x &\mapsto x^2, \end{aligned}$$

$$\begin{aligned} g : \mathbb{Z}_{>0} &\to \mathbb{R} \\ y &\mapsto \ln(y), \end{aligned}$$

then $g \circ f : \mathbb{Z} \smallsetminus \{0\} \to \mathbb{R}$ is the function $x \mapsto \ln(x^2)$.

Composition of functions is *associative*, meaning that if we have functions

$$f : X \to Y, \quad g : Y \to Z, \quad h : Z \to W,$$

then $(h \circ g) \circ f = h \circ (g \circ f)$ (exercise). Therefore, we usually write simply $h \circ g \circ f$, without the brackets.

Identity function

To any set X, we assign a special function called the *identity map* on X. Denoted $\mathbf{1}_X$, it is the function from X to itself that sends each element x to itself. In symbols,

$$\begin{aligned} \mathbf{1}_X : X &\to X \\ x &\mapsto x. \end{aligned}$$

Injectivity, surjectivity, and bijectivity

A function $f : X \to Y$ is called

- *injective* if, for all $x_1, x_2 \in X$, the equality $f(x_1) = f(x_2)$ implies that $x_1 = x_2$,
- *surjective* if, for all $y \in Y$, there is $x \in X$ such that $f(x) = y$,
- *bijective* if, for all $y \in Y$, there is a *unique* $x \in X$ such that $f(x) = y$.

Note that a function is bijective if and only if it is both injective and surjective.

Another way to characterize surjectivity is by reference to the image. Specifically, $f : X \to Y$ is surjective if and only if its image is the whole codomain of f, i.e., $\text{Image}(f) = Y$.

Example. The function $f : \mathbb{R}_{\geq 0} \to \mathbb{R}$ that sends x to x^2 is injective, because if x_1 and x_2 are non-negative real numbers whose squares are equal, then $x_1 = x_2$. However, f is not surjective, because there are real numbers in the codomain, $\mathbb{R}$, that are not in the image, such as -1.

Example. The function $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ that sends x to x^2 is not injective. The domain is all of $\mathbb{R}$ now, and there exist distinct elements x_1, x_2 of the domain such that $g(x_1) = g(x_2)$, such as -1 and 1. However, because we have restricted the codomain to be the set of non-negative real numbers, g is surjective: For every non-negative real number y, there is $x \in \mathbb{R}$ such that $x^2 = y$.

Proposition 2.1. *Let X and Y be non-empty sets and $f : X \to Y$ a function.*

(i) f is injective if and only if there is $g : Y \to X$ such that $g \circ f = \mathbf{1}_X$.

(ii) f is surjective if and only if there is $h : Y \to X$ such that $f \circ h = \mathbf{1}_Y$.

(iii) f is bijective if and only if there is $g : Y \to X$ such that $g \circ f = \mathbf{1}_X$ and $f \circ g = \mathbf{1}_Y$.

For a proof, see Section 1 of the Appendix.

I – 3 Equivalence relations

Imagine that you are in a room in which all the people have blue eyes, brown eyes, green eyes, grey eyes, or hazel eyes. You might choose to group the people together by saying, "Everyone with blue eyes stand here, everyone with brown eyes stand there," and so on. But another way to achieve the same effect would be to say simply, "Group yourself with all people of the same eye colour as you." This is obviously a much more straightforward command to give, and it is more akin to how we group objects together in mathematics.

To make this idea formal, we introduce the notion of a *binary relation* on a set. If A is a set, then a binary relation on A is a subset R of $A \times A$. If $(a, b) \in R$, we write $a \sim_R b$, or just $a \sim b$ if there is no risk of ambiguity.

Example. An example of a binary relation on the set $A = \{1, 2, 3\}$ is

$$R = \{(1, 3), (2, 2), (3, 1), (3, 2)\}.$$

Here, $1 \sim_R 3$, $2 \sim_R 2$, $3 \sim_R 1$, and $3 \sim_R 2$.

A binary relation is called an *equivalence relation* if all of the following hold:

(i) For all $a \in A$, $a \sim_R a$. (reflexivity)

(ii) For all $a, b \in A$, if $a \sim_R b$, then $b \sim_R a$. (symmetry)

(iii) For all $a, b, c \in A$, if $a \sim_R b$ and $b \sim_R c$, then $a \sim_R c$. (transitivity)

The binary relation in the previous example is not an equivalence relation. For example, it is not reflexive, because $1 \not\sim_R 1$. However, the following are all equivalence relations on $\{1, 2, 3\}$:

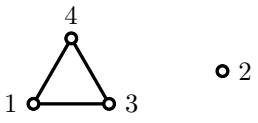
$$R = \{(1, 1), (2, 2), (3, 3)\}$$
$$R = \{(1, 1), (2, 2), (3, 3), (1, 2), (2, 1)\}$$
$$R = \{(1, 1), (2, 2), (3, 3), (1, 2), (2, 1), (1, 3), (3, 1), (2, 3), (3, 2)\}$$

Note how, in each case, reflexivity, symmetry, and transitivity all hold.

An example of an equivalence relation on $\{1, 2, 3, 4\}$ is

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

Since 1, 3, and 4 are all related to one another but 2 is related only to itself, a graphical representation of R would be



(Strictly speaking, the above graphical representation ought to have a "loop" from each number to itself to represent reflexivity, but we have omitted these loops for a cleaner representation.)

It is more common to define a binary relation by declaring a rule to specify when $a \sim b$. For example, the following determine binary relations on $\mathbb{Q}$:

- $a \sim b$ if $a \leq b$.
- $a \sim b$ if $a - b \in \mathbb{Z}$.

Exercise. For each of the above binary relations on $\mathbb{Q}$, determine whether it is an equivalence relation. If any of the three properties fail, which ones?

Equivalence classes

If $\sim$ is an equivalence relation on a set A, then for $a \in A$, the *equivalence class* of a is the set

$$[a] = \{b \in A \mid b \sim a\}.$$

For example, if we define a relation $\sim$ on $\mathbb{Z}$ by specifying that $a \sim b$ if $a - b$ is an integer multiple of 3, then $\sim$ is an equivalence relation (exercise), and there are three equivalence classes:

$$[0] = \{\ldots, -9, -6, -3, 0, 3, 6, 9, \ldots\}$$
$$[1] = \{\ldots, -8, -5, -2, 1, 4, 7, 10, \ldots\}$$
$$[2] = \{\ldots, -7, -4, -1, 2, 5, 8, 11, \ldots\}$$

Note that $[0] = [3] = [6]$, and so on, and similarly for the other classes. Note also how every integer is in exactly one of these equivalence classes. In fact, in general we have the following:

Proposition 3.1. *If $\sim$ is an equivalence relation on a set A, then the equivalence classes partition A, i.e., every element of A is in exactly one equivalence class. Two equivalence classes $[a]$ and $[b]$ are equal if and only if $a \sim b$.*

Proof. Let $a \in A$. Then $a \in [a]$, because $a \sim a$ by reflexivity.

Now suppose that $a \in [b]$, so

$$a \sim b \quad \text{(by definition)}, \tag{3.1}$$
$$b \sim a \quad \text{(by symmetry)}. \tag{3.2}$$

We show that $[a] = [b]$. If $c \in [a]$, then $c \sim a$, so $c \sim b$ by (3.1) together with transitivity, and so $c \in [b]$. Thus, $[a] \subseteq [b]$. On the other hand, if $d \in [b]$, then $d \sim b$, so $d \sim a$ by (3.2) together with transitivity, and so $d \in [a]$.

We have simultaneously proven the second claim of the proposition. □

I – 4 Induction

Suppose that the houses along one side of a certain street have the property that, whenever one house has a red door, so does the house immediately to the right of it. Suppose also that the left-most house has a red door. What can we conclude?

If you realized that all of the houses along that side of the street must have a red door, then you have understood the main idea behind induction.

We now make the notion of induction more formal.

First form of induction

Let $n_0 \in \mathbb{Z}$, and for each $n \geq n_0$, let $P(n)$ be a statement depending on n. Assume that

(i) $P(n_0)$ is true, and

(ii) for all $n \geq n_0$, if $P(n)$ is true, then $P(n+1)$ is true.

Then $P(n)$ is true for all $n \geq n_0$.

Thus, in a proof that uses this first form of induction, we show that the case $n = n_0$ holds (this is often called the *base case*), then assume the statement holds for some $n \geq n_0$ (this is the *inductive hypothesis*), and then show that the statement $P(n+1)$ holds.

Example. We will show by induction that

$$\sum_{k=1}^{n} \frac{1}{k(k+1)} = \frac{n}{n+1} \tag{4.1}$$

for all $n \geq 1$. Here, $n_0 = 1$, and the statement $P(n)$ is the assertion that the equality in (4.1) holds for that n.

The base case, $n = 1$, is true because both sides of the sum take the same value, $1/2$, in this case.

Now let $n \geq 1$, and assume that (4.1) holds for this n (the inductive hypothesis). We wish to show that (4.1) holds with n replaced by $n+1$. But

$$\begin{aligned}
\sum_{k=1}^{n+1} \frac{1}{k(k+1)} &= \sum_{k=1}^{n} \frac{1}{k(k+1)} + \frac{1}{(n+1)(n+2)} \\
&= \frac{n}{n+1} + \frac{1}{(n+1)(n+2)} \quad \text{by the inductive hypothesis} \\
&= \frac{n(n+2)+1}{(n+1)(n+2)} \\
&= \frac{n+1}{n+2},
\end{aligned}$$

as desired. By induction, we are done.

The next example involves the *factorial* of a non-negative integer n, defined by

$$n! = \prod_{k=1}^{n} k.$$

Note that 0!, being the empty product, is equal to 1. The next few factorials are $1! = 1$, $2! = 2$, $3! = 6$, and $4! = 24$.

Example. We show by induction that $2^n < n!$ for all $n \geq 4$. The base case, $n = 4$, is true because $2^4 = 16 < 24 = 4!$. Now let $n \geq 4$ and assume that $2^n < n!$ (the inductive hypothesis). Then

$$\begin{aligned} 2^{n+1} &= 2 \cdot 2^n \\ &< 2 \cdot n! \quad \text{by the inductive hypothesis} \\ &< (n+1)n! \\ &= (n+1)!. \end{aligned}$$

The induction is complete.

Second form of induction

Let $n_0 \in \mathbb{Z}$, and for each $n \geq n_0$, let $P(n)$ be a statement depending on n. Assume that

(i) $P(n_0)$ is true, and

(ii) for all $n \geq n_0$, if $P(k)$ is true for all $k \in \{n_0, \ldots, n\}$, then $P(n+1)$ is true.

Then $P(n)$ is true for all $n \geq n_0$.

In a proof by induction that uses this second form, we show that the case $n = n_0$ holds (the base case), then let $n \geq n_0$ and assume that $P(k)$ holds for all $k \in \{n_0, \ldots, n\}$ (this is the inductive hypothesis in the second form of induction), and then show that the statement $P(n+1)$ holds.

Remark. The two forms of induction are equivalent, i.e., each implies the other, as we prove in Section 2 of the Appendix.

Example. We show, using the second form of induction, that for every integer $n \geq 12$, there are $x, y \in \mathbb{Z}_{\geq 0}$ such that $n = 4x + 5y$. The base case, $n = 12$, is true because $12 = 4 \cdot 3 + 5 \cdot 0$. Now let $n \geq 12$, and assume that, for all $k \in \{12, \ldots, n\}$, there exist $x, y \in \mathbb{Z}_{\geq 0}$ such that $k = 4x + 5y$ (the inductive hypothesis). We want to show that $n + 1$ can be thus expressed as well. The trick is to subtract 4 from $n + 1$, apply the inductive hypothesis to $(n + 1) - 4 = n - 3$, and then add 4 back on.

To do this, we note first that $13 = 2 \cdot 4 + 1 \cdot 5$, $14 = 1 \cdot 4 + 2 \cdot 5$, and $15 = 0 \cdot 4 + 3 \cdot 5$, so we may assume that $n \geq 15$. Then $n - 3 \in \{12, \ldots, n\}$, so by the inductive hypothesis, there are $x, y \in \mathbb{Z}_{\geq 0}$ such that $n - 3 = 4x + 5y$. Therefore,

$$n + 1 = (n - 3) + 4 = 4x + 5y + 4 = 4(x + 1) + 5y,$$

as desired. By induction, we are done.

(II) The Integers

II – 1 Division and greatest common divisors

If $a, b \in \mathbb{Z}$, then b is said to *divide* a if there is $c \in \mathbb{Z}$ such that $a = bc$. In this case, we write $b \,|\, a$.

Remark. The integer 0 is divisible by every integer, even by 0, because if $b \in \mathbb{Z}$, then $0 = b \cdot 0$.

Proposition 1.1 (Division with remainder)**.** *Let $a, b \in \mathbb{Z}$, and assume that $b \neq 0$. Then there are unique $q, r \in \mathbb{Z}$ with $0 \leq r < |b|$ such that $a = qb + r$.*

Proof. For existence, we first prove by induction that, for each $a \geq 0$, there are $q, r \in \mathbb{Z}$ such that $0 \leq r < |b|$ and $a = qb + r$. The case $a = 0$ holds, because $0 = 0 \cdot b + 0$. Now let $a \geq 0$, and assume that there are $q, r \in \mathbb{Z}$ such that $0 \leq r < |b|$ and $a = qb + r$. We have two cases to consider.

Case (i): $r \leq |b| - 2$. Then $a + 1 = qb + (r + 1)$, and $r + 1$ is still less than $|b|$.

Case (ii): $r = |b| - 1$. In this case,

$$a + 1 = qb + r + 1 = qb + |b| = \begin{cases} (q+1)b & \text{if } b > 0 \\ (q-1)b & \text{if } b < 0. \end{cases}$$

By induction, we have settled existence when $a \geq 0$.

For $a < 0$, we observe that $-a - 1 \geq 0$, so by the above, there are $q, r \in \mathbb{Z}$ with $0 \leq r < |b|$ such that $-a - 1 = qb + r$. Hence,

$$a = -qb - r - 1 = -qb - |b| + |b| - r - 1 = -\left(q + \frac{|b|}{b}\right)b + (|b| - r - 1).$$

Note that $|b|/b = \pm 1$ and is therefore an integer. Further, because $0 \leq r < |b|$, we have $0 \leq |b| - r - 1 < |b|$, so we are done.

It remains to prove uniqueness. Suppose that $q_1 b + r_1 = q_2 b + r_2$, where q_1, q_2, r_1, r_2 are integers with $0 \leq r_1, r_2 < |b|$. Then $(q_1 - q_2)b = r_2 - r_1$, so b divides $r_2 - r_1$. But $|r_2 - r_1| < |b|$, so $r_2 - r_1 = 0$, i.e., $r_1 = r_2$. Hence, $q_1 b = q_2 b$, so $q_1 = q_2$. □

The integers q and r in Proposition 1.1 are called the *quotient* and *remainder* of the division.

Greatest common divisors

If $a, b \in \mathbb{Z}$, a *common divisor* of a and b is an integer that divides both a and b. For example, the common divisors of 15 and 35 are $1, -1, 5, -5$.

We define a *greatest common divisor* of a and b to be a positive common divisor of a and b that is divisible by all common divisors.

Theorem 1.2 (G.C.D. Theorem)**.** *Let $a, b \in \mathbb{Z}$, not both zero.*

(i) A greatest common divisor of a and b exists, and it is unique. We denote it $\gcd(a, b)$.

(ii) There are $m, n \in \mathbb{Z}$ such that $\gcd(a, b) = ma + nb$.

Proof. Let $I = \{ma + nb \mid m, n \in \mathbb{Z}\}$. Observe that I is closed under subtraction, which is to say that if $k, l \in I$, then so is $k - l$. Also, it is closed under multiplication by any integer, i.e., if $k \in I$ and $t \in \mathbb{Z}$, then $tk \in I$.

Because a and b are not both zero, I contains positive elements, so let d be the least positive element of I. We claim that $I = d\mathbb{Z}$, the set of all integer multiples of d. One direction is clear: Because $d \in I$ and I is closed under multiplication by integers, it follows that $d\mathbb{Z} \subseteq I$. Conversely, take any $c \in I$. By Proposition 1.1, there are $q, r \in \mathbb{Z}$ with $0 \leq r < d$ such that $c = qd + r$. Then $r = c - qd$, which is in I because I is closed under multiplication by integers and under subtraction. Thus, r is a non-negative element of I that is less than d, so because d is the least positive element of I, we conclude that $r = 0$. Hence, $c = qd \in d\mathbb{Z}$. We have proven our claim that $I = d\mathbb{Z}$, that is,

$$\{ma + nb \mid m, n \in \mathbb{Z}\} = d\mathbb{Z}.$$

Now, $a, b \in I$, so $a, b \in d\mathbb{Z}$, which is to say that d divides both. Thus, d is a positive common divisor of a and b. Further, $d \in I$, so there are $m, n \in \mathbb{Z}$ such that $d = ma + nb$. Therefore, if e divides both a and b, then e divides d. Thus, d is a greatest common divisor of a and b.

Finally, to prove uniqueness, suppose that d and d' are both greatest common divisors of a and b. Then $d' \mid d$ and $d \mid d'$, so $d = d'$ because they are both positive. □

Remark. Although the greatest common divisor of a and b is unique, the integers m and n appearing in part (ii) of Theorem 1.2 are not. Indeed, if $a' = a/d \in \mathbb{Z}$ and $b' = b/d \in \mathbb{Z}$, then for all $k \in \mathbb{Z}$,

$$ma + nb = ma + kda'b' - kda'b' + nb = (m + kb')a + (n - ka')b.$$

Example. Here are some greatest common divisors:

a	b	$\gcd(a, b)$
4	6	2
−4	6	2
0	11	11
5	10	5
15	35	5
126	147	21

We will see next an efficient algorithm for finding greatest common divisors.

II – 2 The Euclidean algorithm

Let a_0 and a_1 be positive integers whose greatest common divisor we wish to find. By division with remainder, we may write

$$a_0 = q_0 a_1 + a_2 \tag{2.1}$$

with $q_0, a_2 \in \mathbb{Z}$ and $0 \leq a_2 < a_1$. From (2.1), we see that any common divisor of a_0 and a_1 is a common divisor of a_1 and a_2, and vice versa. Therefore, $\gcd(a_0, a_1) = \gcd(a_1, a_2)$. This may not, at first sight, appear to be a gain, because we seem only to have replaced one pair of integers with another. But note that $a_2 < a_1$, and if $a_2 \neq 0$, we may repeat the above to write

$$a_1 = q_1 a_2 + a_3$$

with $0 \leq a_3 < a_2$, and so on, until eventually some a_n must be zero. The final step would therefore say $a_{n-2} = q_{n-2} a_{n-1}$, so

$$\gcd(a_0, a_1) = \gcd(a_1, a_2) = \gcd(a_2, a_3) = \cdots = \gcd(a_{n-2}, a_{n-1}) = a_{n-1}.$$

Example. Let us find $\gcd(142, 66)$ by the method just described:

$$142 = 2 \cdot 66 + 10 \tag{2.2}$$

$$66 = 6 \cdot 10 + 6 \tag{2.3}$$

$$10 = 1 \cdot 6 + 4 \tag{2.4}$$

$$6 = 1 \cdot 4 + 2 \tag{2.5}$$

Because $2 \mid 4$, we see that $\gcd(142, 66) = 2$. It is not necessary to write out the final step, $4 = 2 \cdot 2 + 0$.

Expressing $\gcd(a, b)$ in terms of a and b

Recall from Theorem 1.2 that there are $m, n \in \mathbb{Z}$ such that $\gcd(a, b) = ma + nb$. It is possible to extend the algorithm above to find such integers. We illustrate how by continuing with the previous example, in which $a = 142$ and $b = 66$:

$$\begin{aligned}
2 &= 6 - 1 \cdot 4 \quad \text{by (2.5)} \\
&= 6 - 1 \cdot (10 - 1 \cdot 6) \quad \text{because } 4 = 10 - 1 \cdot 6 \text{ by (2.4)} \\
&= 2 \cdot 6 - 10 \\
&= 2(66 - 6 \cdot 10) - 10 \quad \text{because } 6 = 66 - 6 \cdot 10 \text{ by (2.3)} \\
&= 2 \cdot 66 - 13 \cdot 10 \\
&= 2 \cdot 66 - 13(142 - 2 \cdot 66) \quad \text{because } 10 = 142 - 2 \cdot 66 \text{ by (2.2)} \\
&= 28 \cdot 66 - 13 \cdot 142.
\end{aligned}$$

The term *Euclidean algorithm* may refer to the process of finding the greatest common divisor or to the process of expressing the greatest common divisor in terms of the original integers, but commonly the term means both stages combined.

Example. Use the Euclidean algorithm to find the greatest common divisor of $a = 9464$ and $b = 825$ and to express it in the form $ma + nb$ with $m, n \in \mathbb{Z}$.

Solution:

$$9464 = 11 \cdot 825 + 389 \tag{2.6}$$
$$825 = 2 \cdot 389 + 47 \tag{2.7}$$
$$389 = 8 \cdot 47 + 13 \tag{2.8}$$
$$47 = 3 \cdot 13 + 8 \tag{2.9}$$
$$13 = 8 + 5 \tag{2.10}$$
$$8 = 5 + 3 \tag{2.11}$$
$$5 = 3 + 2 \tag{2.12}$$
$$3 = 2 + 1. \tag{2.13}$$

Therefore, $\gcd(9464, 825) = 1$, and

$$\begin{aligned}
1 &= 3 - 2 \quad \text{by (2.13)} \\
&= 3 - (5 - 3) \quad \text{by (2.12)} \\
&= 2 \cdot 3 - 5 \\
&= 2(8 - 5) - 5 \quad \text{by (2.11)} \\
&= 2 \cdot 8 - 3 \cdot 5 \\
&= 2 \cdot 8 - 3(13 - 8) \quad \text{by (2.10)} \\
&= 5 \cdot 8 - 3 \cdot 13 \\
&= 5(47 - 3 \cdot 13) - 3 \cdot 13 \quad \text{by (2.9)} \\
&= 5 \cdot 47 - 18 \cdot 13 \\
&= 5 \cdot 47 - 18(389 - 8 \cdot 47) \quad \text{by (2.8)} \\
&= 149 \cdot 47 - 18 \cdot 389 \\
&= 149(825 - 2 \cdot 389) - 18 \cdot 389 \quad \text{by (2.7)} \\
&= 149 \cdot 825 - 316 \cdot 389 \\
&= 149 \cdot 825 - 316(9464 - 11 \cdot 825) \quad \text{by (2.6)} \\
&= 3625 \cdot 825 - 316 \cdot 9464.
\end{aligned}$$

II – 3 The Fundamental Theorem of Arithmetic

Essential to the study of the integers is the notion of a *prime number*, defined to be an integer greater than 1 whose only positive divisors are 1 and itself. The sequence of primes begins $2, 3, 5, 7, 11, 13, 17, 19, 23, \ldots$.

Lemma 3.1 (Unique-factorization lemma)**.** *Let p be a prime. If $a, b \in \mathbb{Z}$ and $p \mid ab$, then $p \mid a$ or $p \mid b$.*

Proof. Assume that $p \mid ab$, and let $d = \gcd(p, a)$. Being a positive divisor of p, d is either 1 or p. If $d = p$, then $p \mid a$. Otherwise, $d = 1$, so by Theorem 1.2 there exist $m, n \in \mathbb{Z}$ such that $1 = mp + na$, and so $b = mpb + nab$. But because $p \mid ab$, there is $c \in \mathbb{Z}$ such that $ab = pc$, so $b = mpb + npc = p(mb + nc)$, and so $p \mid b$. □

Theorem 3.2 (Fundamental Theorem of Arithmetic)**.** *Every positive integer can be factorized into a product of primes, and the factorization is unique up to the order of the prime factors. (We allow 1 to be considered the empty product of primes.)*

Proof. We prove the existence of a prime factorization of an integer $n \geq 1$ by induction. The base case, $n = 1$, holds because 1 is trivially a product of primes, being the empty product. Now let $n \geq 1$, and assume that every positive integer $k \leq n$ is a product of primes. (Note that we are using the second form of induction, here.) If $n+1$ is prime, we are done. Otherwise, $n + 1 = ab$ where a and b are integers greater than 1 and less than $n + 1$. Therefore, by the inductive hypothesis, each of a and b is a product of primes, so the same is true of $ab = n + 1$. The induction is complete.

For uniqueness, suppose that

$$p_1 p_2 \cdots p_r = q_1 q_2 \cdots q_s \tag{3.1}$$

where the p_i and q_j are primes. We may assume that $r \leq s$. Because $p_1 \mid q_1 q_2 \cdots q_s$, it follows from Lemma 3.1 that $p_1 \mid q_j$ for some j. Reordering if necessary, we may assume that $p_1 \mid q_1$. But p_1 and q_1 are prime, so $p_1 = q_1$. Hence, cancelling $p_1 = q_1$ from both sides of (3.1) gives

$$p_2 p_3 \cdots p_r = q_2 q_3 \cdots q_s.$$

Continuing in this way (an induction argument would make this precise), we arrive at $1 = q_{r+1} q_{r+2} \cdots q_s$, but because no prime can divide 1, the product on the right is the empty product. Therefore, $r = s$ and $p_i = q_i$ for all i (after a reordering of the q_j if necessary). □

Corollary 3.3. *There are infinitely many primes.*

Proof. This proof is attributed to Euclid. Suppose there were only finitely many primes, $p_1, \ldots, p_r$. Let $N = p_1 p_2 \cdots p_r + 1$. Because N is an integer greater than 1, it has a prime divisor by the existence part of Theorem 3.2. This prime divisor must be p_i for some i. But then p_i divides both N and $p_1 p_2 \cdots p_r = N - 1$, so it divides $N - (N - 1) = 1$, a contradiction. Therefore, there are infinitely many primes. □

Two integers are said to be *coprime* if the only positive integer dividing both is 1. The following exercise concerns two very important properties of a pair of coprime integers.

Exercise. Let a, b, c be integers such that a and b are coprime. By Theorem 1.2, there are $m, n \in \mathbb{Z}$ such that $1 = ma + nb$. Use this fact to prove the following two statements:

(i) If $a \mid bc$, then $a \mid c$.

(ii) If $a \mid c$ and $b \mid c$, then $ab \mid c$.

Show by counterexample that both statements would be false if the assertion that a and b be coprime were removed.

(III) Modular Arithmetic

III – 1 Congruences

Let n be a positive integer, called the *modulus*. If $a, b \in \mathbb{Z}$, we say that a is *congruent* to b mod n if $n \mid a - b$. In this case, we write $a \equiv b \bmod n$.

For example, $13 \equiv 28 \bmod 5$, because $5 \mid 13 - 28$. Also, $-7 \equiv 41 \bmod 16$, because $16 \mid -7 - 41$.

Remark. Another way to think about congruence is via remainders. Specifically, $a \equiv b \bmod n$ if and only if a and b have the same remainder when divided by n. For example, 13 and 28 both have the remainder 3 when divided by 5. However, this point of view does not generalize well to other situations we will encounter in this course, so we do not emphasize it. Rather, we will always use the definition of congruence given above.

For a fixed modulus n, the relation of congruence mod n defines an equivalence relation on $\mathbb{Z}$:

(i) For all $a \in \mathbb{Z}$, $a \equiv a \bmod n$.

(ii) For all $a, b \in \mathbb{Z}$, if $a \equiv b \bmod n$, then $b \equiv a \bmod n$.

(iii) For all $a, b, c \in \mathbb{Z}$, if $a \equiv b \bmod n$ and $b \equiv c \bmod n$, then $a \equiv c \bmod n$.

We leave it as a short exercise to prove the above three properties.

Proposition 1.1. *Let n be a positive integer, and let a, a', b, b' be integers such that $a \equiv a' \bmod n$ and $b \equiv b' \bmod n$. Then*

$$\begin{aligned} a + b &\equiv a' + b' && \bmod n \\ ab &\equiv a'b' && \bmod n \end{aligned}$$

Proof. By assumption, $a - a' = kn$ and $b - b' = ln$ for some $k, l \in \mathbb{Z}$. Then

$$(a + b) - (a' + b') = (a - a') + (b - b') = kn + ln = n(k + l),$$

so $a + b \equiv a' + b' \bmod n$. Further,

$$\begin{aligned} ab - a'b' &= ab - a'b + a'b - a'b' \\ &= (a - a')b + a'(b - b') \\ &= knb + a'ln \\ &= n(kb + a'l), \end{aligned}$$

so $ab \equiv a'b' \bmod n$. □

Example. Compute $26\,011 + 598\,312 \bmod 100$ and $26\,011 \cdot 598\,312 \bmod 100$.

Solution: By Proposition 1.1,

$$26\,011 + 598\,312 \equiv 11 + 12 \qquad \bmod 100$$

$$\equiv 23 \qquad \text{mod } 100$$

$$\text{and} \quad 26\,011 \cdot 598\,312 \equiv 11 \cdot 12 \qquad \text{mod } 100$$
$$= 132$$
$$\equiv 32 \qquad \text{mod } 100$$

Example. Show that every positive integer is congruent to the sum of its decimal digits mod 3 and also mod 9. Deduce that $9 \mid 123\,456\,789$.

Solution: Let a be a positive integer, and write $a = \sum_{k=0}^{r} 10^k d_k$ where $d_k \in \{0, \ldots, 9\}$ for all k. We show that $a \equiv \sum_{k=0}^{r} d_k \bmod 9$, from which it follows immediately that the congruence holds mod 3 as well, since $3 \mid 9$.

Because $10 \equiv 1 \bmod 9$, we have $10^k \equiv 1 \bmod 9$ for all $k \geq 0$ by Proposition 1.1, and hence the proposition used again gives

$$a \equiv \sum_{k=0}^{r} 1 \cdot d_k \qquad \text{mod } 9$$
$$= \sum_{k=0}^{r} d_k.$$

For the second part of the question, note that the digits are, from left to right,

$$1,\ 2,\ 3,\ 3+1,\ 3+2,\ 3+3,\ 6+1,\ 6+2,\ 6+3,$$

so the sum of the digits is $3(1+2+3)+3(0+3+6) = 3\cdot 6+3\cdot 9 \equiv 0 \bmod 9$. Alternatively, we could use the fact that

$$1 + 2 + 3 + \cdots + 9 = \frac{1}{2} \cdot 9 \cdot 10 = 9 \cdot 5 \equiv 0 \bmod 9.$$

III – 2 Solving single congruences

Inverses in modular arithmetic

Recall that two integers are said to be coprime if the only positive integer dividing both is 1. In modular arithmetic, if a is coprime to the modulus n, then an integer b such that $ba \equiv 1 \bmod n$ is called an *inverse* of $a \bmod n$. We find such a b via the Euclidean algorithm, as in Section II – 2.

Remark. When $ba \equiv 1 \bmod n$, we sometimes call b *the* inverse of $a \bmod n$, because although b is not unique, it is unique mod n in the following sense: If $ba \equiv 1 \bmod n$ and $b'a \equiv 1 \bmod n$, then $b \equiv b' \bmod n$ (exercise).

Example. Find an inverse to 13 mod 54.

Solution:

$$54 = 4 \cdot 13 + 2$$
$$13 = 6 \cdot 2 + 1,$$

so

$$1 = 13 - 6 \cdot 2 = 13 - 6(54 - 4 \cdot 13) = 25 \cdot 13 - 6 \cdot 54.$$

Thus, $25 \cdot 13 \equiv 1 \bmod 54$, so 25 is inverse to 13 mod 54.

Coefficient of an unknown in a congruence

Suppose we have a term of the form ax in a congruence, where a is known and x is an unknown to be solved for. The following proposition describes how we can isolate x in two common situations.

Proposition 2.1. *Let n be a positive integer, and let $x, y \in \mathbb{Z}$.*

(i) If a is a positive integer, then $ax \equiv ay \bmod an$ if and only if $x \equiv y \bmod n$.

(ii) If a is an integer coprime to n, then $ax \equiv ay \bmod n$ if and only if $x \equiv y \bmod n$.

Proof. (i) $ax \equiv ay \bmod an$ if and only if $an \mid ax - ay$, if and only if $an \mid a(x - y)$, if and only if $n \mid x - y$, if and only if $x \equiv y \bmod n$.

(ii) One direction is immediate: If $x \equiv y \bmod n$, then $ax \equiv ay \bmod n$ by Proposition 1.1. This does not use the fact that a and n are coprime. The converse, however, does. Indeed, if a and n are coprime, then by Theorem 1.2 in Section II, there are $b, c \in \mathbb{Z}$ such that $1 = ba + cn$, so $ba \equiv 1 \bmod n$. Hence, multiplying both sides of the congruence $ax \equiv ay \bmod n$ by b to obtain $bax \equiv bay \bmod n$, we see that $1 \cdot x \equiv 1 \cdot y \bmod n$, i.e., $x \equiv y \bmod n$. □

Example. Solve the congruence $28x \equiv 14 \bmod 77$.

Solution: By part (i) of Proposition 2.1, $28x \equiv 14 \bmod 77$ if and only if $4x \equiv 2 \bmod 11$. Now, 4 is coprime to the modulus 11, and we can find an inverse via the Euclidean algorithm:

$$\begin{aligned} 11 &= 2 \cdot 4 + 3 \\ 4 &= 3 + 1, \end{aligned}$$

so

$$1 = 4 - 3 = 4 - (11 - 2 \cdot 4) = 3 \cdot 4 - 11,$$

and so 3 is inverse to 4 mod 11, i.e., $3 \cdot 4 \equiv 1 \bmod 11$. Hence,

$$\begin{aligned} & 4x \equiv 2 && \bmod 11 && \\ \iff & 3 \cdot 4x \equiv 3 \cdot 2 && \bmod 11 && \text{by part (ii) of Proposition 2.1} \\ \iff & x \equiv 6 && \bmod 11. && \end{aligned}$$

III – 3 Solving simultaneous congruences

Let m and n be positive integers, and let a and b be any integers. Consider the system

$$\begin{aligned} x &\equiv a \mod m \\ x &\equiv b \mod n \end{aligned}$$

of simultaneous congruences. In general, this system need not have a solution. Consider, for example, $m = 6$, $n = 10$, $a = 1$, and $b = 2$. If $x \equiv 1 \bmod 6$, then x must be odd, but if $x \equiv 2 \bmod 10$, then x must be even. Therefore, there is no simultaneous solution.

However, if the moduli m and n are coprime, then there is a solution. Indeed, in that case, there are integers $s, t \in \mathbb{Z}$ such that

$$1 = sm + tn, \tag{3.1}$$

so letting $x_0 = atn + bsm$, we see that

$$\begin{aligned} x_0 &\equiv atn && \mod m \\ &= a - asm && \text{by (3.1)} \\ &\equiv a && \mod m, \end{aligned}$$

$$\begin{aligned} \text{and} \quad x_0 &\equiv bsm && \mod n \\ &= b - btn && \text{by (3.1) again} \\ &\equiv b && \mod n. \end{aligned}$$

In fact, if x_0 is any simultaneous solution (not necessarily even the one constructed above), then an integer x is a simultaneous solution if and only if $x \equiv x_0 \bmod mn$. To see this, note first that if $x \equiv x_0 \bmod mn$, then certainly $x \equiv x_0 \equiv a \bmod m$, and $x \equiv x_0 \equiv b \bmod n$. Conversely, if $x \equiv a \bmod m$ and $x \equiv b \bmod n$, then $x \equiv x_0 \bmod m$ and $x \equiv x_0 \bmod n$, so both m and n divide $x - x_0$. Therefore, because m and n are coprime, part (ii) of the exercise in Section II–3 shows that $mn \mid x - x_0$, i.e., $x \equiv x_0 \bmod mn$.

Example. Find all integers x such that $x \equiv 7 \bmod 16$ and $x \equiv 2 \bmod 21$.

Solution: Via the Euclidean algorithm, we find that $1 = 4 \cdot 16 - 3 \cdot 21$, so a solution is $x_0 = 7(-3) \cdot 21 + 2 \cdot 4 \cdot 16 = -313$. Therefore, a given integer x is a solution if and only if $x \equiv -313 \bmod 16 \cdot 21$, if and only if $x \equiv -313 \bmod 336$, if and only if $x \equiv 23 \bmod 336$.

Systems of more than two congruences

If we have more than two congruences, where the moduli are pairwise coprime, then we may solve them by iterating the method above, as we illustrate in the next example.

Example. Solve the following simultaneous congruences:

$$x \equiv 8 \mod 11 \tag{3.2}$$

$$x \equiv 3 \mod 14 \tag{3.3}$$

$$x \equiv 23 \mod 39 \tag{3.4}$$

Solution: Let us solve congruences (3.2) and (3.3) first. The Euclidean algorithm yields

$$1 = -5 \cdot 11 + 4 \cdot 14,$$

so the first two congruences are equivalent to

$$\begin{aligned} & x \equiv 8 \cdot 4 \cdot 14 + 3(-5) \cdot 11 && \mod 11 \cdot 14 \\ & \quad = 283, \\ \text{i.e.,} \quad & x \equiv 129 && \mod 154. \end{aligned} \tag{3.5}$$

Now we solve (3.5) and (3.4). Using the Euclidean algorithm again, we obtain

$$1 = 19 \cdot 154 - 75 \cdot 39,$$

so (3.5) and (3.4) are equivalent to

$$\begin{aligned} & x \equiv 129(-75) \cdot 39 + 23 \cdot 19 \cdot 154 && \mod 154 \cdot 39 \\ & \quad = -310\,027, \\ \text{i.e.,} \quad & x \equiv 2285 && \mod 6006. \end{aligned}$$

Thus, the solutions to the original three congruences (3.2)–(3.4) are the integers x satisfying $x \equiv 2285 \mod 6006$.

Let us summarize the above ideas in a theorem, called the Chinese Remainder Theorem.

Theorem 3.1 (Chinese Remainder Theorem)**.** *Let $n_1, \ldots, n_r$ be pairwise coprime positive integers, and let $a_1, \ldots, a_r$ be any integers. Then the simultaneous congruences*

$$\begin{aligned} x &\equiv a_1 && \mod n_1 \\ &\vdots \\ x &\equiv a_r && \mod n_r \end{aligned}$$

have a solution. Further, if x_0 is any solution, then the set of all solutions consists of those integers x such that $x \equiv x_0 \mod n_1 n_2 \ldots n_r$.

Proof. We proved the case $r = 2$ when we discussed systems of two congruences. The general case follows by induction on $r \geq 2$. □

(IV) Rings

IV – 1 Binary operations and the definition of a ring

A *binary operation* on a set A is a function $* : A \times A \to A$. Usually, we write $a * b$ instead of $*(a, b)$.

Example. If $A = \{1, 2, 3\}$, then an example of a binary operation on A is

$$\begin{aligned} * : A \times A &\to A \\ (1,1) &\mapsto 3 \\ (1,2) &\mapsto 3 \\ (1,3) &\mapsto 2 \\ (2,1) &\mapsto 1 \\ (2,2) &\mapsto 1 \\ (2,3) &\mapsto 3 \\ (3,1) &\mapsto 2 \\ (3,2) &\mapsto 3 \\ (3,3) &\mapsto 1 \end{aligned}$$

According to this operation, $1 * 2 = 3$, $1 * 3 = 2$, $2 * 1 = 1$, $2 * 2 = 1$, and $3 * 2 = 3$, for example.

The set $\mathbb{Z}$ of integers has two especially important binary operations:

$$\begin{aligned} + : \mathbb{Z} \times \mathbb{Z} &\to \mathbb{Z} \\ \times : \mathbb{Z} \times \mathbb{Z} &\to \mathbb{Z} \end{aligned}$$

Note that the multiplication operation $\times$ on $\mathbb{Z}$ is often written as a dot instead, as in $3 \cdot 5 = 15$. In fact, this is the notation we have been using throughout.

Definition of a ring

A *ring* is a non-empty set R together with two binary operations,

$$\begin{aligned} + : R \times R &\to R \\ \cdot : R \times R &\to R, \end{aligned}$$

satisfying all of the following:

(i) *Axioms of addition*

(A1) For all $a, b, c \in R$, $(a + b) + c = a + (b + c)$. (associativity of addition)

(A2) For all $a, b \in R$, $a + b = b + a$. (commutativity of addition)

(A3) There exists $0 \in R$ such that, for all $a \in R$, $0 + a = a$. (existence of a neutral element for addition)

(A4) For all $a \in R$, there exists $b \in R$ such that $b + a = 0$. (existence of additive inverses)

(ii) *Associativity of multiplication*

- For all $a, b, c \in R$, $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.

(iii) *Distributivity*

- For all $a, b, c \in R$,

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

Conventions on notation

(i) $a \cdot b + c \cdot d$ will be taken to mean $(a \cdot b) + (c \cdot d)$, not $a \cdot (b + c) \cdot d$. Similarly, $a + b \cdot c + d$ will be taken to mean $a + (b \cdot c) + d$, not $(a + b) \cdot (c + d)$.

(ii) We will often omit the dot for multiplication: ab instead of $a \cdot b$.

(iii) If n is a positive integer and $a \in R$, we define

$$na = \underbrace{a + \cdots + a}_{n \text{ times}}$$
$$a^n = \underbrace{a \cdots a}_{n \text{ times}}$$

(or, more precisely, we define na recursively by $1a = a$ and $(n+1)a = na + a$, and similarly for a^n). Then for all positive integers n, m and all $a, b \in R$,

$$n(a + b) = na + nb \tag{1.1}$$
$$(n + m)a = na + ma \tag{1.2}$$
$$(nm)a = n(ma) \tag{1.3}$$
$$a^{n+m} = a^n a^m \tag{1.4}$$
$$a^{nm} = (a^n)^m \tag{1.5}$$

Prove these equalities for practice using the recursive definitions of na and a^n.

IV – 2 First properties of rings

Proposition 2.1. *If R is a ring, the element 0 in axiom (A3) is unique.*

Proof. Suppose that both 0 and $0'$ satisfy the axiom, i.e.,

$$0 + a = a \tag{2.1}$$

$$0' + a = a \tag{2.2}$$

for all $a \in R$. Then

$$\begin{aligned} 0' &= 0 + 0' \quad \text{by (2.1) with } a = 0' \\ &= 0' + 0 \quad \text{by commutativity} \\ &= 0 \quad \text{by (2.2) with } a = 0. \end{aligned}$$

□

The unique element 0 satisfying axiom (A3) is called the *zero element* of R. It is often denoted 0_R, but when there is no risk of ambiguity, we often write 0 instead.

Proposition 2.2. *If R is a ring, then for each $a \in R$, the element $b \in R$ appearing in axiom (A4), i.e., such that $b + a = 0_R$, is unique.*

Proof. Let $a \in R$, and suppose that $b, b' \in R$ satisfy $b + a = 0_R$ and $b' + a = 0_R$. Note, of course, that $a + b' = 0_R$ as well by commutativity of addition. Then

$$\begin{aligned} b' &= 0_R + b' \\ &= (b + a) + b' \\ &= b + (a + b') \quad \text{(associativity of addition)} \\ &= b + 0_R \\ &= b. \end{aligned}$$

□

For a given $a \in R$, the unique $b \in R$ such that $b + a = 0_R$ is called the *additive inverse* of a and is denoted $-a$. Thus, $(-a) + a = a + (-a) = 0_R$. If $x, y \in R$, then $x - y$ is defined to be $x + (-y)$.

Proposition 2.3. *If $a \in R$ and 0_R is the zero element of R, then $0_R \cdot a = a \cdot 0_R = 0_R$.*

Proof. Because $0_R = 0_R + 0_R$,

$$0_R \cdot a = (0_R + 0_R) \cdot a = 0_R \cdot a + 0_R \cdot a$$

by distributivity. Hence, adding $-(0_R \cdot a)$ to both sides of the equation $0_R \cdot a = 0_R \cdot a + 0_R \cdot a$ yields

$$\begin{aligned} 0_R \cdot a - 0_R \cdot a &= (0_R \cdot a + 0_R \cdot a) - 0_R \cdot a, \\ \text{i.e.,} \quad 0_R &= 0_R \cdot a. \end{aligned}$$

The equality $a \cdot 0_R = 0_R$ is proven similarly. □

Proposition 2.4. *If $a, b, c \in R$, then $a \cdot (b - c) = a \cdot b - a \cdot c$ and $(b - c) \cdot a = b \cdot a - c \cdot a$.*

Proof.

$$\begin{aligned} a \cdot (b - c) &= a \cdot (b - c) + a \cdot c - a \cdot c \\ &= a \cdot ((b - c) + c) - a \cdot c \quad \text{by distributivity over addition} \\ &= a \cdot b - a \cdot c. \end{aligned}$$

The other equality is proven similarly. □

Adding a ring element to itself a negative number of times

Let R be a ring. We have already seen the definition of na when $a \in R$ and n is a positive integer. Now that we have introduced additive inverses, we may even define na when $n < 0$. Namely, if $a \in R$ and n is a negative integer, then $-n > 0$ and we define

$$na = -(-n)a,$$

that is, na is defined to be the additive inverse of $(-n)a$. For example, $(-4)a = -(4a) = -(a + a + a + a)$. Note that na is not to be considered a multiplication; we are not multiplying n and a together in the ring. Indeed, there is no reason that R should contain integers at all in general.

Finally, for completeness, we define $0a = 0_R$, where $a \in R$, 0_R is the zero element of R, and 0 is the integer zero. To avoid any confusion, it is worth emphasizing at this point that

$$\begin{aligned} 0_R \cdot a = 0_R \quad &\text{by Proposition 2.3,} \\ 0a = 0_R \quad &\text{simply by definition.} \end{aligned}$$

Exercise. Show that for all integers n and all $a \in R$, one has $(n + 1)a = na + a$ and $(n - 1)a = na - a$.

Exercise. Show that (1.1)–(1.3) hold for all integers n, m.

IV – 3 Commutative and unital rings

Many of the rings we study in this course will have one or both of the following common properties.

Commutative rings

A ring R is called *commutative* if its multiplication operation is commutative, i.e., if $a \cdot b = b \cdot a$ for all $a, b \in R$.

Unital rings

A ring R is called *unital* if there is an element $1 \in R$ such that, for all $a \in R$,

$$1 \cdot a = a \cdot 1 = a.$$

Proposition 3.1. *If R is a unital ring, then the element 1 satisfying $1 \cdot a = a \cdot 1 = a$ for all $a \in R$ is unique.*

Proof. Suppose $1, 1' \in R$ satisfy $1 \cdot a = a \cdot 1 = a$ and $1' \cdot a = a \cdot 1' = a$ for all $a \in R$. Then $1' = 1 \cdot 1' = 1$, the first equality because 1 has the property in question, and the second because $1'$ does. □

The unique element 1 satisfying $1 \cdot a = a \cdot 1 = a$ for all $a \in R$ is called the *identity* element of R, or the *multiplicative identity element*. Therefore, a unital ring is sometimes also called a *ring with identity*.

We often denote the identity element in a unital ring by 1_R.

Caution. Many books include the existence of a multiplicative identity in the very definition of a ring. When you are reading a book that discusses rings, make sure to check how rings are defined so that you know whether the author is assuming the existence of a multiplicative identity or not.

If a is an element of a unital ring R, we define $a^0 = 1_R$ no matter what a is, even when $a = 0_R$.

IV – 4 First examples of rings

Let us look at some examples of rings.

Example. The set of integers, $\mathbb{Z}$, together with the usual operations of addition of multiplication, is a ring. It is commutative and unital.

Example. The set of rational numbers, $\mathbb{Q}$, is a ring. Again, it is commutative and unital. The same goes for the set $\mathbb{R}$ of real numbers.

Example. If n is a positive integer, then the set $M_n(\mathbb{R})$ of $n \times n$ matrices with real entries forms a ring, in which the two operations are the usual addition and multiplication of $n \times n$ matrices. The ring $M_n(\mathbb{R})$ is unital, having the $n \times n$ identity matrix as its identity element. However, when $n \geq 2$, $M_n(\mathbb{R})$ is not commutative. For example, in the 2×2 case, we may consider the matrices

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

which satisfy

$$AB = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{but} \quad BA = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

For $n \geq 3$, we may consider matrices of the form

$$A' = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}, \quad B' = \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix},$$

where A and B are as above and the zeroes represent blocks of appropriate sizes. Then $A'B' = B'$, but $B'A' = 0$.

Example. Let $\mathcal{F}$ denote the set of functions $f : \mathbb{R} \to \mathbb{R}$. If we define operations of addition and multiplication on $\mathcal{F}$ by

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

then $\mathcal{F}$ is a ring. Further, it is commutative and unital, the identity element being the constant function $x \mapsto 1$. The commutativity of $\mathcal{F}$ follows from the commutativity of $\mathbb{R}$: If $f, g \in \mathcal{F}$, then for all $x \in \mathbb{R}$,

$$(f \cdot g)(x) = f(x)g(x) = g(x)f(x) = (g \cdot f)(x),$$

so $f \cdot g = g \cdot f$.

Example. If X is a set, we define the *power set* of X, $P(X)$, to be the set of subsets of X. That is,

$$P(X) = \{A \mid A \subseteq X\}.$$

We may define operations of addition and multiplication on $P(X)$ by

$$A + B = (A \smallsetminus B) \cup (B \smallsetminus A) = (A \cup B) \smallsetminus (A \cap B)$$
$$A \cdot B = A \cap B$$

With these operations, $P(X)$ is a commutative ring with identity, the identity element being X. Notice that the zero element of $P(X)$ is $\emptyset$, for if $A \in P(X)$, then $\emptyset + A = (A \cup \emptyset) \smallsetminus (A \cap \emptyset) = A \smallsetminus \emptyset = A$.

For example, if $X = \{1, 2, 3\}$, then

$$P(X) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}, \{3\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$

We have

$$\{1\} + \{2\} = \{1, 2\}, \quad \{1\} \cdot \{2\} = \emptyset, \quad \{1, 2\} + \{2, 3\} = \{1, 3\}, \quad \{1, 2\} \cdot \{2, 3\} = \{2\}.$$

Example. Let n be a positive integer. We denote by $\mathrm{End}(\mathbb{R}^n)$ the set of linear transformations $f : \mathbb{R}^n \to \mathbb{R}^n$. (Recall that a linear transformation satisfies $f(\mathbf{u} + \mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v})$ and $f(c\mathbf{u}) = cf(\mathbf{u})$ for all vectors $\mathbf{u}, \mathbf{v}$ and all $c \in \mathbb{R}$.) We define addition and multiplication of linear transformations by

$$(f + g)(\mathbf{u}) = f(\mathbf{u}) + g(\mathbf{u}) \quad \text{for all } \mathbf{u} \in \mathbb{R}^n$$
$$f \cdot g = f \circ g$$

Then $\mathrm{End}(\mathbb{R}^n)$ is a unital ring, the identity element being the identity linear transformation $\mathbf{u} \mapsto \mathbf{u}$. However, when $n \geq 2$, $\mathrm{End}(\mathbb{R}^n)$ is not commutative (exercise).

Let us verify one of the distributivity properties for $\mathrm{End}(\mathbb{R}^n)$, namely, $f \cdot (g + h) = f \cdot g + f \cdot h$. If $\mathbf{u} \in \mathbb{R}^n$, then

$$\begin{aligned}
(f \cdot (g + h))(\mathbf{u}) &= (f \circ (g + h))(\mathbf{u}) && \text{by definition of multiplication} \\
&= f((g + h)(\mathbf{u})) && \text{by definition of composition} \\
&= f(g(\mathbf{u}) + h(\mathbf{u})) && \text{by definition of addition} \\
&= f(g(\mathbf{u})) + f(h(\mathbf{u})) && \text{because } f \text{ is linear} \\
&= (f \circ g)(\mathbf{u}) + (f \circ h)(\mathbf{u}) && \text{by definition of composition} \\
&= ((f \circ g) + (f \circ h))(\mathbf{u}) && \text{by definition of addition} \\
&= (f \cdot g + f \cdot h)(\mathbf{u}) && \text{by definition of multiplication,}
\end{aligned}$$

so $f \cdot (g + h) = f \cdot g + f \cdot h$, as desired.

Caution. Compare the addition and multiplication operations of $\mathrm{End}(\mathbb{R})$ with those of $\mathcal{F}$. Note that addition is defined the same way in both rings, but their multiplication operations are different from each other.

IV – 5 Polynomials and sequences

Polynomial rings

Let R be a commutative unital ring. A polynomial with coefficients in R is an expression of the form

$$a_n x^n + \cdots + a_1 x + a_0 \tag{5.1}$$

where n is some non-negative integer and the coefficients a_i are elements of the ring R. The set of all such polynomials is denoted $R[x]$.

At first, the powers of x could just be symbols without any particular meaning. We could even just represent the polynomial in (5.1) by its vector of coefficients, $(a_n, \ldots, a_0)$. However, the powers of x play an important role once we turn the set of polynomials into a ring. Let us consider multiplication first, because that is the more interesting operation on polynomials. If $f(x) = a_m x^m + \cdots + a_0$ and $g(x) = b_n x^n + \cdots + b_0$ are polynomials in $R[x]$, then their product $f(x)g(x)$ is defined to be the polynomial

$$f(x)g(x) = \sum_{k=0}^{m+n} \left(\sum_{l=0}^{k} a_l b_{k-l} \right) x^k.$$

Example. If $f(x) = 3x^2 - 4x + 5$ and $g(x) = 4x^3 + 7x^2 + 2x - 3$ are in $\mathbb{Z}[x]$, then the coefficient of, say, x^2 in $f(x)g(x)$ is

$$(3)(-3) + (-4)(2) + (5)(7) = 18.$$

Similarly, the coefficient of x^4 is $(3)(7) + (-4)(4) = 5$. Written out in full,

$$f(x)g(x) = 12x^5 + 5x^4 - 2x^3 + 18x^2 + 22x - 15.$$

Note that, the way multiplication is defined, $x^m x^n = x^{m+n}$, as one would expect.

Addition in $R[x]$ is more straightforward: we just add polynomials coefficient by coefficient. For example, if $f(x) = 3x^2 - 4x + 5$ and $g(x) = 4x^3 + 7x^2 + 2x - 3$ again, then

$$f(x) + g(x) = (0+4)x^3 + (3+7)x^2 + (-4+2)x + 5 + (-3) = 4x^3 + 10x^2 - 2x + 2.$$

With the above operations of addition and multiplication, $R[x]$ forms a commutative unital ring. Its identity element is the polynomial 1.

Remark.

- Two polynomials in $R[x]$ are equal if and only if their coefficients are equal.
- The *degree* $\deg(f)$ of a non-zero polynomial f is the largest n such that the coefficient of x^n in f is non-zero. The zero polynomial is defined to have degree $-\infty$ in this course. For all $f, g \in R[x]$, $\deg(f+g) \leq \max(\deg(f), \deg(g))$, and if R is $\mathbb{Z}$, $\mathbb{Q}$, or $\mathbb{R}$, then $\deg(fg) = \deg(f) + \deg(g)$. We adopt the conventions that $-\infty \leq n$ and $-\infty + n = -\infty$ for all $n \in \mathbb{Z}_{\geq 0} \cup \{-\infty\}$.

- The *leading coefficient* of a non-zero polynomial is the coefficient of the highest power of x occurring in the polynomial.

Evaluation of polynomials

Let R again be a commutative unital ring. If $f(x) = a_n x^n + \cdots + a_0 \in R[x]$ and $b \in R$, then the *evaluation* of f at b, denoted $f(b)$, is the element

$$a_n b^n + \cdots + a_1 b + a_0$$

of R. For a fixed b,

$$(f+g)(b) = f(b) + g(b)$$
$$(fg)(b) = f(b)g(b)$$

for all $f, g \in R[x]$.

Rings of sequences

Let R be any ring. We let $\mathcal{S}(R)$ denote the set of all sequences

$$(a_n)_{n\geq 0} = (a_0, a_1, a_2, \ldots)$$

with entries $a_n \in R$. Addition of sequences is defined componentwise, i.e.,

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

as is multiplication:

$$(a_0, a_1, a_2, \ldots) \cdot (b_0, b_1, b_2, \ldots) = (a_0 \cdot b_0, a_1 \cdot b_1, a_2 \cdot b_2, \ldots).$$

In more succinct notation, we have

$$(a_n)_{n\geq 0} + (b_n)_{n\geq 0} = (a_n + b_n)_{n\geq 0},$$
$$(a_n)_{n\geq 0} \cdot (b_n)_{n\geq 0} = (a_n \cdot b_n)_{n\geq 0}.$$

With addition and multiplication defined this way, $\mathcal{S}(R)$ is a ring. It is commutative if and only if R is, and it is unital if and only if R is. If R does have an identity, then the identity in $\mathcal{S}(R)$ is $(1_R)_{n\geq 0} = (1_R, 1_R, 1_R, \ldots)$.

Note that, in this course, all sequences start with the index $n = 0$. With this convention made, we choose to omit the ≥ 0 in the notation, and write simply $(a_n)_n$ for the sequence $(a_0, a_1, a_2, \ldots)$.

A sequence may have any entries whatsoever. However, some sequences may be given by expressions in the indexing variable n, such as those in the next example.

Example. The following are all sequences in $\mathcal{S}(\mathbb{Z})$:

$$(n)_n = (0, 1, 2, 3, \ldots)$$
$$((n+1)^2)_n = (1, 4, 9, 16, \ldots)$$
$$(\cos(\tfrac{n\pi}{2}))_n = (1, 0, -1, 0, 1, 0, -1, 0, \ldots)$$

IV – 6 Units

Let R be a unital ring. An element $a \in R$ is called a *unit* if there are $b, c \in R$ such that $ba = ac = 1_R$. The set of units in R is denoted $R^\times$.

Proposition 6.1. *If $a \in R^\times$, then the elements $b, c \in R$ such that $ba = ac = 1_R$ are unique and are equal to each other.*

Proof. $b = b \cdot 1_R = b(ac) = (ba)c = 1_R \cdot c = c$. □

We call the unique $b \in R$ such that $ba = ab = 1_R$ the *inverse* of a (or, for emphasis, the *multiplicative inverse*). It is denoted a^{-1}. Thus, $a^{-1}a = aa^{-1} = 1_R$.

Cancellation of units

If $a \in R^\times$ and $b, c \in R$, then

$$\begin{aligned} ab = ac \Rightarrow b = c, \\ \text{and} \quad ba = ca \Rightarrow b = c. \end{aligned}$$

Indeed, if $ab = ac$, then $a^{-1}ab = a^{-1}ac$, so $1_R \cdot b = 1_R \cdot c$, i.e., $b = c$. The second equality is proven similarly.

Units in commutative rings

If R is commutative, then $a \in R$ is a unit if and only if there is $b \in R$ such that $ba = 1_R$. That is, we do not need to check separately that $ab = 1_R$, because $ab = ba$.

Examples of units

(i) $\mathbb{Z}^\times = \{1, -1\}$. That is, the only units in the ring of integers are 1 and -1.

(ii) $\mathbb{Q}^\times = \mathbb{Q} \smallsetminus \{0\}$. That is, every non-zero rational number is a unit in $\mathbb{Q}$. Similarly, $\mathbb{R}^\times = \mathbb{R} \smallsetminus \{0\}$.

(iii) If $f \in \mathcal{F}$, then f is a unit if and only if $f(x)$ is non-zero for all $x \in \mathbb{R}$. Thus, the function $f \in \mathcal{F}$ defined by $f(x) = x^2 + 1$ is a unit in $\mathcal{F}$, and its inverse (as an element of $\mathcal{F}$) is the function $g \in \mathcal{F}$ defined by $g(x) = 1/(x^2 + 1)$. Indeed, $(f \cdot g)(x) = f(x)g(x) = (x^2 + 1)/(x^2 + 1) = 1$ for all $x \in \mathbb{R}$. By contrast, the function $h(x) = (x - 1)^3$ is not a unit, because $h(1) = 0$.

(iv) The units in the ring $M_n(\mathbb{R})$ are the invertible $n \times n$ matrices, or equivalently, the ones with non-zero determinant. The set of invertible $n \times n$ matrices is denoted $GL_n(\mathbb{R})$. For example,

$$\begin{pmatrix} 2 & 1 \\ 3 & 8 \end{pmatrix} \in GL_2(\mathbb{R}), \quad \begin{pmatrix} 2 & 1 \\ 6 & 3 \end{pmatrix} \notin GL_2(\mathbb{R}).$$

IV – 7 The ring $\mathbb{Z}/n\mathbb{Z}$

Let n be a positive integer. Recall from Section III – 1 that, for $a, b \in \mathbb{Z}$, $a \equiv b \bmod n$ if and only if $n \mid a - b$. This is the very definition of congruence. We also saw that congruence mod n defines an equivalence relation on $\mathbb{Z}$. The equivalence classes with respect to congruence are called *residue classes*, and we will denote the residue class of a mod n by $[a]_n$, or simply $[a]$ if there is no risk of ambiguity. Thus,

$$[a]_n = \{b \in \mathbb{Z} \mid b \equiv a \bmod n\} = \{b \in \mathbb{Z} \mid n \text{ divides } b - a\}.$$

Example. The residue classes mod 5 are

$$\begin{aligned}
[0] &= \{\ldots, -5, 0, 5, 10, \ldots\} \\
[1] &= \{\ldots, -4, 1, 6, 11, \ldots\} \\
[2] &= \{\ldots, -3, 2, 7, 12, \ldots\} \\
[3] &= \{\ldots, -2, 3, 8, 13, \ldots\} \\
[4] &= \{\ldots, -1, 4, 9, 14, \ldots\}
\end{aligned}$$

If n is a positive integer, the set of residue classes mod n is denoted $\mathbb{Z}/n\mathbb{Z}$. In general,

$$\mathbb{Z}/n\mathbb{Z} = \{[0], [1], [2], \ldots, [n-1]\}.$$

Note that each element of $\mathbb{Z}/n\mathbb{Z}$ has many representatives. For example, in $\mathbb{Z}/5\mathbb{Z}$, $[2] = [7] = [-3]$, and so on, that is, the residue class $[2]$ is represented not only by 2, but also by 7, by -3, and in fact by $[b]$ for any $b \equiv 2 \bmod 5$. We will need to take care over this phenomenon when we define binary operations on $\mathbb{Z}/n\mathbb{Z}$.

Operations of addition and multiplication

We define an operation of addition on $\mathbb{Z}/n\mathbb{Z}$ by

$$[a] + [b] = [a + b].$$

Let us check that this operation is well defined, i.e., is independent of the chosen representatives a and b. If $[a] = [a']$ and $[b] = [b']$, then $a \equiv a' \bmod n$ and $b \equiv b' \bmod n$, so $a + b \equiv a' + b' \bmod n$ by Proposition 1.1 in Section III, and so $[a + b] = [a' + b']$, as required.

Example. Let $x = [3]$ and $y = [4]$ in $\mathbb{Z}/5\mathbb{Z}$. We will calculate $x + y$ in two ways. First,

$$x + y = [3] + [4] = [3 + 4] = [7].$$

But $x = [18]$ and $y = [-21]$, so

$$x + y = [18] + [-21] = [18 - 21] = [-3].$$

Of course, $[7] = [-3]$, so everything works out.

Similarly, we define an operation of multiplication on $\mathbb{Z}/n\mathbb{Z}$ by

$$[a] \cdot [b] = [ab].$$

Again, this is well defined by Proposition 1.1 in Section III. That is, if $[a] = [a']$ and $[b] = [b']$, then $[a + b] = [a' + b']$.

Example. If $x = [3]$ and $y = [5]$ in $\mathbb{Z}/7\mathbb{Z}$, then

$$x \cdot y = [3] \cdot [5] = [15] = [1].$$

We would arrive at the same result if we instead used the representatives 10 and 19 for x and y:

$$x \cdot y = [10] \cdot [19] = [190] = [1],$$

the last equality holding because $190 - 1 = 189 = 27 \cdot 7$.

With the above operations, $\mathbb{Z}/n\mathbb{Z}$ is a commutative unital ring. The zero element is $[0]$, and the identity is $[1]$. We leave it as an exercise to verify the ring axioms, but let us verify associativity of multiplication by way of example. If $a, b, c \in \mathbb{Z}$, then

$$\begin{aligned}
([a] \cdot [b]) \cdot [c] &= [ab] \cdot [c] && \text{by definition of multiplication} \\
&= [(ab)c] && \text{by definition of multiplication again} \\
&= [a(bc)] && \text{by associativity of multiplication in } \mathbb{Z} \\
&= [a] \cdot [bc] && \text{by definition of multiplication again} \\
&= [a] \cdot ([b] \cdot [c]) && \text{by definition of multiplication once more.}
\end{aligned}$$

You will find that the other ring axioms reduce in a similar way to the corresponding axioms for the ring $\mathbb{Z}$. This phenomenon will reappear when we study quotient rings.

Units in $\mathbb{Z}/n\mathbb{Z}$

Proposition 7.1. *Let n be a positive integer and $a \in \mathbb{Z}$. Then $[a] \in (\mathbb{Z}/n\mathbb{Z})^\times$ if and only if $\gcd(a, n) = 1$, i.e., if and only if a and n are coprime.*

Proof. Assume that $[a] \in (\mathbb{Z}/n\mathbb{Z})^\times$, i.e., there is $b \in \mathbb{Z}$ such that $[b] \cdot [a] = [1]$. Then $n \mid ba - 1$, so there is $k \in \mathbb{Z}$ such that $ba - 1 = kn$, i.e., $1 = ba - kn$. Therefore, any common divisor of a and n divides 1, so a and n are coprime.

Conversely, assume that a and n are coprime. Then by Theorem 1.2, there are $b, c \in \mathbb{Z}$ such that $1 = ba + cn$, so $n \mid ba - 1$, and so $[ba] = [1]$, i.e., $[b] \cdot [a] = [1]$. □

Example.

$$(\mathbb{Z}/40\mathbb{Z})^\times = \{[1], [3], [7], [9], [11], [13], [17], [19], [21], [23], [27], [29], [31], [33], [37], [39]\}.$$

Exercise. In $\mathbb{Z}/21\mathbb{Z}$, $[16]$ is a unit because $\gcd(16, 21) = 1$. By finding $[16]^{-1}$, solve the equation $[16]x = [15]$ in $\mathbb{Z}/21\mathbb{Z}$. You may find the Euclidean algorithm useful to find the inverse.

IV – 8 Subrings and products

Let R be a ring. A *subring* of R is a subset of R that is itself a subring with respect to the same operations of addition and multiplication in R. For example, $\mathbb{Z}$ is a subring of $\mathbb{Q}$.

Proposition 8.1. *Let R be a ring and S a subset of R. Then S is a subring of R if and only if all of the following hold:*

(i) S is non-empty.

(ii) For all $a, b \in S$, $a - b \in S$. (closure under subtraction)

(iii) For all $a, b \in S$, $ab \in S$. (closure under multiplication)

Proof. Let us begin the proof to show some key aspects of it, and leave the rest as an exercise. Assume that S is non-empty and is closed under subtraction. Being non-empty, it contains some element a, so it also contains $a - a = 0_R$ because it is closed under subtraction. Hence, if $c \in S$, then S contains $0_R - c = -c$, again because S is closed under subtraction. Therefore, by closure under subtraction once more, we see that if $b, c \in S$, then S contains $b - (-c) = b + c$. Thus, S is closed under addition, so the addition operation of R restricts to a binary operation on S.

If S is also closed under multiplication, it follows immediately that the multiplication operation on R restricts to a binary operation on S. It is then straightforward to verify that these two operations on S, addition and multiplication, satisfy the ring axioms.

Conversely, if S is a subring of R, then it must be non-empty and closed under subtraction and multiplication. □

Example. Let n be a positive integer, and let $S = \{A \in M_n(\mathbb{R}) \mid A \text{ is upper triangular}\}$. (Recall that a square matrix $A = (a_{i,j})_{i,j}$ is said to be *upper triangular* if $a_{i,j} = 0$ for all $i > j$.) We show that S is a subring of $M_n(\mathbb{R})$. It is non-empty because, for example, the zero $n \times n$ matrix is upper triangular. It is closed under subtraction because if $A = (a_{i,j})$ and $B = (b_{i,j})$ are both upper triangular, meaning that $a_{i,j} = b_{i,j} = 0$ when $i > j$, then $a_{i,j} - b_{i,j} = 0$ when $i > j$. Further, the (i, j)-entry of AB is

$$\sum_{k=1}^{n} a_{i,k} b_{k,j},$$

and if $i > j$, then every $k \in \{1, \ldots, n\}$ satisfies either $k < i$, in which case $a_{i,k} = 0$, or $k \geq i > j$, in which case $b_{k,j} = 0$, so AB is upper triangular.

Example. A function $f : \mathbb{R} \to \mathbb{R}$ is called *bounded* if there is $M > 0$ such that, for all $x \in \mathbb{R}$, $|f(x)| \leq M$. We show that the subset S of $\mathcal{F}$ consisting of the bounded functions is a subring. It is non-empty because the zero function is bounded. Now suppose that f and g are two bounded functions, so there are $M, N > 0$ such that, for all $x \in \mathbb{R}$, $|f(x)| \leq M$ and $|g(x)| \leq N$. Then

$$|(f - g)(x)| = |f(x) - g(x)| \leq |f(x)| + |g(x)| \leq M + N,$$

$$\text{and} \quad |(f \cdot g)(x)| = |f(x)g(x)| = |f(x)|\,|g(x)| \leq MN,$$

so $f - g$ and $f \cdot g$ are both bounded.

Example. The subset $\mathbb{Z}_{\geq 0}$ of $\mathbb{Z}$ consisting of the non-negative integers is not a subring. Indeed, although it is non-empty and is closed under multiplication, it is not closed under subtraction, for 0 and 1 are non-negative but $0 - 1 = -1$ is not.

Example. Let $n \geq 2$ be an integer, and let

$$S = \{A \in M_n(\mathbb{R}) \mid \operatorname{Tr}(A) = 0\}.$$

Then S is not a subring of $M_n(\mathbb{R})$. This time, it is closed under subtraction, but it is not closed under multiplication. For example, if $A \in M_n(\mathbb{R})$ is the diagonal matrix in which the first diagonal entry is 1, the second is -1, and all the rest are zero, then $A \in S$ but $A^2 \notin S$.

Products of rings

If $R_1, \ldots, R_n$ are rings, we may define operations of addition and multiplication on the cartesian product $R_1 \times \cdots \times R_n$ as follows:

$$(a_1, \ldots, a_n) + (b_1, \ldots, b_n) = (a_1 + b_1, \ldots, a_n + b_n),$$
$$(a_1, \ldots, a_n) \cdot (b_1, \ldots, b_n) = (a_1 \cdot b_1, \ldots, a_n \cdot b_n),$$

where each a_i and b_i is in R_i. These operations turn $R_1 \times \cdots \times R_n$ into a ring, called the *product* of $R_1, \ldots, R_n$.

Example.

$$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z} = \{([0]_2, [0]_3), ([0]_2, [1]_3), ([0]_2, [2]_3), ([1]_2, [0]_3), ([1]_2, [1]_3), ([1]_2, [2]_3)\}.$$

Exercise. In $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$, compute $n([1]_2, [1]_3)$ for various $n \in \mathbb{Z}$. What do you notice? What happens if you repeat the exercise with $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$ instead?

Example. In $\mathbb{Z} \times \mathbb{Q} \times (\mathbb{Z}/6\mathbb{Z})$,

$$(3, 1/2, [4]) + (-1, 5, [3]) = (2, 11/2, [1]),$$
$$(3, 1/2, [4]) \cdot (-1, 5, [3]) = (-3, 5/2, [0]).$$

Exercise. Show that if R is a ring, then $\{(r, r) \in R \times R \mid r \in R\}$ is a subring of $R \times R$.

IV – 9 Ideals

Let R be a ring. An *ideal* of R is a non-empty subset I of R such that both of the following hold:

(i) For all $a, b \in I$, $a - b \in I$. (closure under subtraction)

(ii) For all $a \in I$ and all $r \in R$, $ra \in I$ and $ar \in I$. (closure under multiplication by ring elements)

Remark.

- Every ideal is closed under addition. The proof is identical to the first paragraph of the proof of Proposition 8.1.
- If R is commutative, then to check (ii), it is enough to check that $ra \in I$ for all $a \in I$ and all $r \in R$.

Example. Let $I = 2\mathbb{Z} = \{2m \mid m \in \mathbb{Z}\}$, the set of even integers. We show that I is an ideal of $\mathbb{Z}$. It is non-empty, because it contains the integers $2m$ with $m \in \mathbb{Z}$. Next, we take $a, b \in I$, say $a = 2m$ and $b = 2n$ where $m, n \in \mathbb{Z}$. Then $a - b = 2m - 2n = 2(m - n) \in I$, so I is closed under subtraction. Finally, if $a \in I$, say $a = 2m$, and $b \in \mathbb{Z}$, then $ba = b(2m) = 2bm \in I$, so I is closed under multiplication by ring elements.

Example. Let I consist of the elements of $\mathbb{Z} \times \mathbb{Z}$ of the form $(a, 0)$, where $a \in \mathbb{Z}$. Then I is an ideal of $\mathbb{Z} \times \mathbb{Z}$. It is non-empty, because it contains the elements $(a, 0)$. Next, if $(a, 0)$ and $(b, 0)$ are two such elements, then $(a, 0) - (b, 0) = (a - b, 0)$, which is again in I. Finally, if $(a, 0) \in I$ and $(b, c) \in \mathbb{Z} \times \mathbb{Z}$, then $(b, c) \cdot (a, 0) = (ba, 0) \in I$, so I is closed under multiplication by ring elements. Note that $\mathbb{Z} \times \mathbb{Z}$ is commutative, so we do not need to check separately that $(a, 0) \cdot (b, c) \in I$.

Example. Let us see an example of an ideal in a non-commutative ring. Let

$$I = \{A \in M_2(\mathbb{Z}) \mid \text{every entry of } A \text{ is even}\}.$$

(Here, $M_2(\mathbb{Z})$ denotes the ring of 2×2 matrices with integer entries.) By observing that the zero 2×2 matrix has all its entries equal to zero, we see that I is non-empty. Next, if $A, B \in I$, so that $A = 2X$ and $B = 2Y$ for some $X, Y \in M_2(\mathbb{Z})$, then $A - B = 2X - 2Y = 2(X - Y) \in I$, so I is closed under subtraction. Finally, if $A \in I$, say $A = 2X$, and $B \in M_2(\mathbb{Z})$, then $BA = B(2X) = 2BX \in I$, and $AB = 2XB \in I$, so I is closed under multiplication by ring elements. Note how we checked that both BA and AB were in I.

Every ideal is a subring (exercise: use Proposition 8.1), but not every subring is an ideal, as the next example shows.

Example. Recall from Section 8 that the subset S of $\mathcal{F}$ consisting of the bounded functions is a subring of $\mathcal{F}$. However, it is not an ideal. Indeed, if $f : x \mapsto 1$ and $g : x \mapsto x$, then $f \in S$, but $g \cdot f = g \notin S$, so S is not closed under multiplication by ring elements.

Generators and principal ideals

Let R be a unital ring, and let X be a subset of R ($X = \emptyset$ is allowed). We denote by $(X)_R$ the intersection of all ideals containing X. If there is no risk of ambiguity, we may write simply (X) without the subscript. We have the following:

(i) $(X)_R$ is an ideal of R. It is called the *ideal generated by* X.

(ii) $(X)_R$ is in fact the smallest ideal of R containing X, in the sense that if I is any ideal of R containing X, then $(X)_R \subseteq I$.

(iii) $(\emptyset)_R = \{0\}$, the zero ideal.

(iv) $(1)_R = R$.

Ideals generated by finite sets $X = \{a_1, \ldots, a_n\}$ have a special notation. In this case, we write $(X)_R = (a_1, \ldots, a_n)_R$. An ideal I generated by a single element is called *principal*, i.e., $I = (a)_R$ for some $a \in R$.

The commutative case

If R is a commutative unital ring, then an ideal generated by a finite set has a very simple description. Specifically, if $a_1, \ldots, a_n \in R$, then

$$\begin{aligned}(a_1, \ldots, a_n)_R &= \{a_1b_1 + \cdots + a_nb_n \mid b_1, \ldots, b_n \in R\} \\ &= \{b_1a_1 + \cdots + b_na_n \mid b_1, \ldots, b_n \in R\}.\end{aligned}$$

In particular,

$$\begin{aligned}(a)_R &= \{ab \mid b \in R\} = aR, \\ \text{or equivalently,} \quad (a)_R &= \{ba \mid b \in R\} = Ra.\end{aligned}$$

Example. In $\mathbb{Z}$,

$$\begin{aligned}(15, 35)_{\mathbb{Z}} = \{15a + 35b \mid a, b \in \mathbb{Z}\} &= (\gcd(15, 35))_{\mathbb{Z}} \quad \text{by Theorem 1.2 in Section II} \\ &= (5)_{\mathbb{Z}} = 5\mathbb{Z}.\end{aligned}$$

Example. In contrast to the previous example, we show that in $\mathbb{Z}[x]$ the ideal

$$I = (2, x)_{\mathbb{Z}[x]} = \{2g_1 + xg_2 \mid g_1, g_2 \in \mathbb{Z}[x]\}$$

is not principal. Assume that I is principal, say $I = (f)_{\mathbb{Z}[x]} = f\mathbb{Z}[x]$. Then there are $g, h \in \mathbb{Z}[x]$ such that $2 = fg$ and $x = fh$. The equality $2 = fg$ implies that f is one of the four constant polynomials $1, -1, 2, -2$. If $f = \pm 1$, then $1 \in I$, so $1 = 2g_1 + xg_2$ for some $g_1, g_2 \in \mathbb{Z}[x]$. But this is not possible, because $2g_1 + xg_2$ has an even constant term. Therefore, $f = \pm 2$. But then the equality $x = fh$ says that $x = \pm 2h$, which is again not possible, because the coefficient of x in $2h$ is even. We have thus arrived at a contradiction, so I is not principal.

IV – 10 Quotient rings

Recall that we defined operations of addition and multiplication on $\mathbb{Z}/n\mathbb{Z}$ by $[a]+[b] = [a+b]$ and $[a]\cdot[b] = [ab]$. We now generalize this idea.

Let R be a ring and I an ideal of R. Define a binary relation $\sim_I$ on R by declaring that $a \sim_I b$ if $a - b \in I$. We leave it as an exercise to show that this is an equivalence relation.

If $a \in R$, the equivalence class of a is

$$\begin{aligned}\{b \in R \mid b \sim_I a\} &= \{b \in R \mid b - a \in I\} \\ &= \{a + c \mid c \in I\},\end{aligned}$$

which we write $a + I$.

The equivalence class $a + I$ is called a *coset* of I. It is a sort of "shift" of the ideal I by the element a.

Example. Let $I = (x)_{\mathbb{Z}[x]} = x\mathbb{Z}[x] = \{xg \mid g \in \mathbb{Z}[x]\}$. Let us examine the coset $3 + I$, for example. It consists of the polynomials of the form $3 + f$ with $f \in I$, i.e., $3 + xg$ where $g \in \mathbb{Z}[x]$. Put another way, $3 + I$ consists of the polynomials in $\mathbb{Z}[x]$ that have constant term equal to 3.

Proposition 10.1. *If I is an ideal of R, and $a, b \in R$, then $a + I = b + I$ if and only if $a - b \in I$.*

Proof. Because a coset is an equivalence class, $a + I = b + I$ if and only if $a \sim_I b$ by Proposition 3.1 in Section I, if and only if $a - b \in I$ by definition of the relation $\sim_I$. □

Example. In light of Proposition 10.1, we may easily describe the coset $(x^2+2x+4)+I$ of $\mathbb{Z}[x]$, where again $I = x\mathbb{Z}[x]$. Indeed, $(x^2 + 2x + 4) - 4 = x^2 + 2x \in I$, so

$$(x^2 + 2x + 4) + I = 4 + I = \{4 + f \mid f \in I\} = \{4 + xg \mid g \in \mathbb{Z}[x]\}.$$

Thus, $(x^2 + 2x + 4) + I$ consists of the polynomials in $\mathbb{Z}[x]$ having constant term equal to 4.

Example. Let $I = \{f \in \mathcal{F} \mid f(n\pi) = 0 \text{ for all } n \in \mathbb{Z}\}$, an ideal of $\mathcal{F}$ (exercise). If $g \in \mathcal{F}$ is defined by $g(x) = \cos(2x)$, show that the coset $g+I$ is equal to $f+I$ for some constant function $f \in \mathcal{F}$.

Solution: If $n \in \mathbb{Z}$, then $g(n\pi) = \cos(2n\pi) = 1$. Therefore, if f is the constant function $x \mapsto 1$, we have

$$(g - f)(n\pi) = g(n\pi) - f(n\pi) = 1 - 1 = 0 \quad \text{for all } n \in \mathbb{Z},$$

so $g - f \in I$, that is, $g + I = f + I$.

Operations of addition and multiplication

If R is a ring and I an ideal of R, we denote by R/I the set of cosets $a+I$ of I. That is,

$$R/I = \{a + I \mid a \in R\}.$$

We may define operations of addition and multiplication on R/I as follows:

$$(a + I) + (b + I) = (a + b) + I$$
$$(a + I) \cdot (b + I) = (ab) + I$$

These operations are well defined in the sense that, if $a + I = a' + I$ and $b + I = b' + I$, then $(a + b) + I = (a' + b') + I$ and $(ab) + I = (a'b') + I$. The proof is identical to that in the case of $\mathbb{Z}/n\mathbb{Z}$ in Section 7.

These operations make R/I a ring, called the *quotient* of R by I. The zero element of R/I is $0_R + I = I$. If R is unital, then so is R/I, the identity being $1_R + I$.

Remark. If $a \in R$, then $-(a+I) = (-a)+I$, because $((-a)+I)+(a+I) = (-a+a)+I = 0_R + I$.

Example. Let $I = (x^2 + 1)_{\mathbb{R}[x]} = (x^2 + 1)\mathbb{R}[x] = \{(x^2 + 1)g \mid g \in \mathbb{R}[x]\}$, an ideal in $\mathbb{R}[x]$. The quotient ring $S = \mathbb{R}[x]/I$ has identity $1_S = 1 + I$, where 1 is the constant polynomial. We show that -1_S is square in S:

$$\begin{aligned}(x + I)^2 = x^2 + I = (-1) + I \quad &\text{because } x^2 - (-1) = x^2 + 1 \in I \\ &= -(1 + I) = -1_S.\end{aligned}$$

In fact, this is one way to construct the complex numbers, which we will come to later.

Example. Let $I = \{(a_n)_n \in \mathcal{S}(\mathbb{Z}) \mid 2 \text{ divides } a_n \text{ for all } n \geq 0\}$, an ideal of $\mathcal{S}(\mathbb{Z})$. Show that $x^2 = x$ for all $x \in \mathcal{S}(\mathbb{Z})/I$.

Solution: Observe that $a^2 \equiv a \bmod 2$ for all $a \in \mathbb{Z}$, i.e., $a^2 - a$ is even. Therefore, if $\alpha = (a_n)_n \in \mathcal{S}(\mathbb{Z})$, then $\alpha^2 - \alpha = (a_n^2 - a_n)_n \in I$, i.e., $\alpha^2 + I = \alpha + I$, i.e., $(\alpha + I)^2 = \alpha + I$.

Example. Let $I = \{h \in \mathcal{F} \mid h(n\pi/2) = 0 \text{ for all } n \in \mathbb{Z}\}$, an ideal of $\mathcal{F}$, and define $f, g \in \mathcal{F}$ by $f(x) = \sin(x)$ and $g(x) = \cos(x)$. Show that $(f + I)(g + I) = 0_S$ in $S = \mathcal{F}/I$, but that neither $f + I$ nor $g + I$ is equal to 0_S.

Solution: If n is an even integer, then $\sin(n\pi/2) = 0$, and if n is odd, then $\cos(n\pi/2) = 0$. Thus, $(fg)(n\pi/2) = 0$ for all $n \in \mathbb{Z}$, so $fg \in I$. (Alternatively, we may use the trigonometric identity $\sin(x)\cos(x) = \frac{1}{2}\sin(2x)$.) Hence,

$$(f + I)(g + I) = (fg) + I = I = 0_S.$$

To see that $f + I \neq 0_S$, we have only to show that $f \notin I$, but this is clear because $f(\pi/2) = 1 \neq 0$. Similarly, $g \notin I$ because $g(0) = 1 \neq 0$.

(V) Ring Homomorphisms

V – 1 Definition and first examples of ring homomorphisms

Let R and S be rings. A *ring homomorphism* is a map $\varphi : R \to S$ such that the following both hold for all $a, b \in R$:

(i) $\varphi(a+b) = \varphi(a) + \varphi(b)$. ($\varphi$ respects addition)

(ii) $\varphi(ab) = \varphi(a)\varphi(b)$. ($\varphi$ respects multiplication)

Example. Fix a positive integer n, and define

$$\begin{array}{rcl} \varphi : \mathbb{Z} & \to & \mathbb{Z}/n\mathbb{Z} \\ a & \mapsto & [a]. \end{array}$$

Then φ is a ring homomorphism, because if $a, b \in \mathbb{Z}$,

$$\begin{aligned} \varphi(a+b) &= [a+b] \quad \text{by definition of } \varphi \\ &= [a] + [b] \quad \text{by definition of addition in } \mathbb{Z}/n\mathbb{Z} \\ &= \varphi(a) + \varphi(b) \quad \text{by definition of } \varphi \text{ again,} \end{aligned}$$

$$\begin{aligned} \text{and} \quad \varphi(ab) &= [ab] \quad \text{by definition of } \varphi \\ &= [a][b] \quad \text{by definition of multiplication in } \mathbb{Z}/n\mathbb{Z} \\ &= \varphi(a)\varphi(b) \quad \text{by definition of } \varphi \text{ again.} \end{aligned}$$

Example. Let R be the set of upper-triangular 2×2 matrices with integer entries, a subring of $M_2(\mathbb{Z})$. That is, matrices in R take the form $\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}$ with $a, b, c \in \mathbb{Z}$. Define

$$\begin{array}{rcl} \varphi : R & \to & \mathbb{Z} \times \mathbb{Z} \\ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} & \mapsto & (a, c). \end{array}$$

Then φ is a ring homomorphism:

$$\begin{aligned} \varphi\left(\begin{pmatrix} a & b \\ 0 & c \end{pmatrix} + \begin{pmatrix} a' & b' \\ 0 & c' \end{pmatrix}\right) &= \varphi\begin{pmatrix} a+a' & b+b' \\ 0 & c+c' \end{pmatrix} \quad \text{by definition of addition in } R \\ &= (a+a', c+c') \quad \text{by definition of } \varphi \\ &= (a, c) + (a', c') \quad \text{by definition of addition in } \mathbb{Z} \times \mathbb{Z} \\ &= \varphi\begin{pmatrix} a & b \\ 0 & c \end{pmatrix} + \varphi\begin{pmatrix} a' & b' \\ 0 & c' \end{pmatrix} \quad \text{by definition of } \varphi \text{ again,} \end{aligned}$$

$$\begin{aligned} \text{and} \quad \varphi\left(\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}\begin{pmatrix} a' & b' \\ 0 & c' \end{pmatrix}\right) &= \varphi\begin{pmatrix} aa' & ab'+bc' \\ 0 & cc' \end{pmatrix} \quad \text{by def. of multiplication in } R \\ &= (aa', cc') \quad \text{by definition of } \varphi \end{aligned}$$

$$= (a,c)(a',c') \quad \text{by definition of multiplication in } \mathbb{Z}\times\mathbb{Z}$$
$$= \varphi\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}\varphi\begin{pmatrix} a' & b' \\ 0 & c' \end{pmatrix} \quad \text{by definition of } \varphi \text{ again.}$$

Example. Let R be a commutative unital ring, and fix $b \in R$. In Section IV – 5, we stated that for polynomials $f, g \in R[x]$,

$$(f+g)(b) = f(b) + g(b)$$
$$(fg)(b) = f(b)g(b).$$

These equations say that the map

$$\begin{aligned} \varphi_b : R[x] &\to R \\ f &\mapsto f(b) \end{aligned}$$

is a ring homomorphism.

Example. The map

$$\begin{aligned} \varphi : M_2(\mathbb{Z}) &\to \mathbb{Z} \\ A &\mapsto \mathrm{Tr}(A) \end{aligned}$$

is not a ring homomorphism, because it does not respect multiplication. To see this, note that if I is the 2×2 identity matrix, then

$$\varphi(I^2) = \varphi(I) = \mathrm{Tr}(I) = 2,$$
$$\text{while} \quad \varphi(I)^2 = \mathrm{Tr}(I)^2 = 4.$$

Note, however, that φ does respect addition.

Example. The map

$$\begin{aligned} \varphi : \mathbb{Z} &\to \mathbb{Z} \\ a &\mapsto |a| \end{aligned}$$

is not a ring homomorphism, because it does not respect addition. For example, if $a = 1$ and $b = -1$, then

$$\varphi(a+b) = \varphi(0) = 0,$$
$$\text{while} \quad \varphi(a) + \varphi(b) = 1 + 1 = 2.$$

V – 2 Basic properties of ring homomorphisms

Proposition 2.1. *If $\varphi : R \to S$ is a ring homomorphism, then $\varphi(0_R) = 0_S$.*

Proof. $\varphi(0_R) = \varphi(0_R + 0_R) = \varphi(0_R) + \varphi(0_R)$, so adding $-\varphi(0_R)$ to both sides leaves $0_S = \varphi(0_R)$. □

Proposition 2.2. *Ring homomorphisms respect subtraction, i.e., $\varphi(a-b) = \varphi(a)-\varphi(b)$ for all $a, b \in R$.*

Proof.

$$\begin{aligned}\varphi(a-b) &= \varphi(a-b) + \varphi(b) - \varphi(b) \\ &= \varphi((a-b)+b) - \varphi(b) \quad \text{because } \varphi \text{ respects addition} \\ &= \varphi(a) - \varphi(b).\end{aligned}$$

□

If $\varphi : R \to S$ is a ring homomorphism, we define its *kernel* to be the set

$$\operatorname{Ker}(\varphi) = \{a \in R \mid \varphi(a) = 0_S\}.$$

Proposition 2.3. *If $\varphi : R \to S$ is a ring homomorphism, then $\operatorname{Ker}(\varphi)$ is an ideal of R.*

Proof. Proposition 2.1 shows that $\operatorname{Ker}(\varphi)$ is non-empty. Now suppose that $a, b \in \operatorname{Ker}(\varphi)$. Then

$$\varphi(a-b) \overset{\text{Prop. 2.2}}{=} \varphi(a) - \varphi(b) = 0_S - 0_S = 0_S,$$

so $a - b \in \operatorname{Ker}(\varphi)$, and so $\operatorname{Ker}(\varphi)$ is closed under subtraction. Also, if $a \in \operatorname{Ker}(\varphi)$ and $r \in R$, then

$$\varphi(ra) = \varphi(r)\varphi(a) = \varphi(r) \cdot 0_S = 0_S$$

by Proposition 2.3 in Section IV, so $ra \in \operatorname{Ker}(\varphi)$. Similarly, $ar \in \operatorname{Ker}(\varphi)$. Thus, $\operatorname{Ker}(\varphi)$ is closed under multiplication by ring elements. □

Proposition 2.4. *If $\varphi : R \to S$ is a ring homomorphism, then φ is injective if and only if $\operatorname{Ker}(\varphi) = \{0_R\}$.*

Proof. Assume that φ is injective. If $a \in \operatorname{Ker}(\varphi)$, then $\varphi(a) = 0_S = \varphi(0_R)$, so $a = 0_R$ by injectivity. Conversely, assume that $\operatorname{Ker}(\varphi) = \{0_R\}$, and suppose that $a, b \in R$ satisfy $\varphi(a) = \varphi(b)$. Then

$$\varphi(a-b) = \varphi(a) - \varphi(b) = 0_S,$$

so $a - b \in \operatorname{Ker}(\varphi) = \{0_R\}$, and so $a = b$. □

Examples of kernels

Example. Recall the ring homomorphism

$$\begin{aligned} \varphi : R &\to \mathbb{Z} \times \mathbb{Z} \\ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} &\mapsto (a, c) \end{aligned}$$

where R is the ring of upper-triangular 2×2 matrices with integer entries. A matrix $\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}$ is in $\mathrm{Ker}(\varphi)$ if and only if $(a, c) = (0, 0)$, if and only if $a = c = 0$, so $\mathrm{Ker}(\varphi)$ consists of the matrices of the form $\begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix}$ where $b \in \mathbb{Z}$. Because $\mathrm{Ker}(\varphi)$ contains non-zero matrices (take $b \neq 0$), φ is not injective.

Example. Consider the ring homomorphism

$$\begin{aligned} \varphi : \mathbb{Z} &\to \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z} \\ a &\mapsto ([a]_3, [a]_4). \end{aligned}$$

If $a \in \mathbb{Z}$, then $\varphi(a) = ([0]_3, [0]_4)$ if and only if $[a]_3 = [0]_3$ and $[a]_4 = [0]_4$, if and only if both 3 and 4 divide a, if and only if 12 divides a by part (ii) of the exercise in Section II – 3. (We have used the fact that 3 and 4 are coprime.) Thus, $\mathrm{Ker}(\varphi) = 12\mathbb{Z}$. In particular, because $\mathrm{Ker}(\varphi)$ is not zero, φ is not injective.

Example. Consider the map

$$\begin{aligned} \varphi : \mathbb{Z}[x] &\to \mathcal{S}(\mathbb{Z}) \\ f &\mapsto (f(0), f(1), f(2), \ldots), \end{aligned}$$

a ring homomorphism (exercise). It is a fact that if $f \in \mathbb{Z}[x]$ is a polynomial of degree at most n and there exist $n + 1$ distinct $a \in \mathbb{Z}$ such that $f(a) = 0$, then f is the zero polynomial. Use this fact to show that φ is injective.

Solution: Suppose that $f \in \mathrm{Ker}(\varphi)$. Then $(f(0), f(1), f(2), \ldots) = (0, 0, 0, \ldots)$, so $f(a) = 0$ for all $a \in \mathbb{Z}$. In particular, no matter what degree f has, there are more than $\deg(f)$ integers a for which $f(a) = 0$, so f is the zero polynomial. Thus, $\mathrm{Ker}(\varphi) = \{0\}$, so φ is injective.

Image of a ring homomorphism

Proposition 2.5. *If $\varphi : R \to S$ is a ring homomorphism, then* $\mathrm{Image}(\varphi)$ *is a subring of S.*

Proof. We show that $\mathrm{Image}(S)$ is closed under multiplication and leave the rest as an exercise. Let $s, t \in \mathrm{Image}(\varphi)$, i.e., $s = \varphi(a)$ and $t = \varphi(b)$ for some $a, b \in R$. Then $st = \varphi(a)\varphi(b) = \varphi(ab) \in \mathrm{Image}(\varphi)$. □

Exercise. Find rings R and S and a ring homomorphism $\varphi : R \to S$ such that $\mathrm{Image}(\varphi)$ is not an ideal of S.

V – 3 Ring isomorphisms

A ring homomorphism $\varphi : R \to S$ is called an *isomorphism* if there is a ring homomorphism $\psi : S \to R$ such that $\psi \circ \varphi = \mathbf{1}_R$ and $\varphi \circ \psi = \mathbf{1}_S$. In other words,

$$\begin{aligned} \psi(\varphi(r)) &= r \quad \text{for all } r \in R \\ \text{and} \quad \varphi(\psi(s)) &= s \quad \text{for all } s \in S. \end{aligned}$$

Proposition 3.1. *Let $\varphi : R \to S$ be a ring homomorphism. Then the following are equivalent:*

(i) φ is an isomorphism.

(ii) φ is bijective (i.e., injective and surjective).

(iii) $\mathrm{Ker}(\varphi) = \{0_R\}$ *and* $\mathrm{Image}(\varphi) = S$.

For a proof, see Section 3 of the Appendix.

If there is an isomorphism $\varphi : R \to S$, we say that R is *isomorphic* to S and write $R \cong S$. If this is the case, then the inverse map $\varphi^{-1} : S \to R$ is an isomorphism as well, so $S \cong R$.

Example. Let $S = \{(a, a) \in \mathbb{Z} \times \mathbb{Z}\}$, a subring of $\mathbb{Z} \times \mathbb{Z}$. It consists of the elements in $\mathbb{Z} \times \mathbb{Z}$ of the form (a, a). Define

$$\begin{aligned} \varphi : \mathbb{Z} &\to S \\ a &\mapsto (a, a). \end{aligned}$$

This map is a ring homomorphism, because if $a, b \in \mathbb{Z}$, then

$$\begin{aligned} \varphi(a + b) = (a + b, a + b) = (a, a) + (b, b) = \varphi(a) + \varphi(b) \\ \text{and} \quad \varphi(ab) = (ab, ab) = (a, a)(b, b) = \varphi(a)\varphi(b). \end{aligned}$$

Its kernel consists of the integers a such that $(a, a) = (0, 0)$, i.e., only the integer 0, so it is injective. Further, it is surjective, because given any $(a, a) \in S$, $(a, a) = \varphi(a)$. Thus, φ is an isomorphism by Proposition 3.1.

Example. Let $X = \{1, 2\}$, and define

$$\begin{aligned} \varphi : P(X) &\to \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \\ \emptyset &\mapsto ([0], [0]) \\ \{1\} &\mapsto ([1], [0]) \\ \{2\} &\mapsto ([0], [1]) \\ \{1, 2\} &\mapsto ([1], [1]) \end{aligned}$$

We show that φ is a ring isomorphism. It is clear from the definition that it is bijective, so it remains to show that it respects addition and multiplication. If $A \in P(X)$, then

the first entry of $\varphi(A)$ records whether or not $1 \in A$, while the second entry records whether $2 \in A$. Now, $A + B$ is the "exclusive OR" of A and B, meaning that a given element of X is in $A + B$ if and only if it is in one of A or B but not both. On the other hand, addition in $\mathbb{Z}/2\mathbb{Z}$ has exactly the same property: $x + y = [1]$ if and only if one, but not both, of x and y is equal to $[1]$. Therefore, $\varphi(A + B) = \varphi(A) + \varphi(B)$. The argument for multiplication is similar: $AB = A \cap B$, so an element of X is in AB if and only if it is in both A and B, while in the ring $\mathbb{Z}/2\mathbb{Z}$, $xy = [1]$ if and only if both x and y are equal to $[1]$.

The First Isomorphism Theorem

We show how to construct an isomorphism in a natural way from any given ring homomorphism.

Theorem 3.2 (First Isomorphism Theorem)**.** *Let R and S be rings and $\varphi : R \to S$ a ring homomorphism. There is a well-defined ring homomorphism*

$$\begin{aligned} \overline{\varphi} : R/\operatorname{Ker}(\varphi) &\to \operatorname{Image}(\varphi) \\ a + \operatorname{Ker}(\varphi) &\mapsto \varphi(a), \end{aligned}$$

and $\overline{\varphi}$ is an isomorphism.

Proof. For brevity, let $I = \operatorname{Ker}(\varphi)$. Let us first show that $\overline{\varphi}$ is well defined. If $a + I = a' + I$, then $a - a' \in I = \operatorname{Ker}(\varphi)$, so

$$\varphi(a) = \varphi(a - a' + a') = \varphi(a - a') + \varphi(a') = 0_S + \varphi(a') = \varphi(a').$$

Next, we show that $\overline{\varphi}$ is a ring homomorphism:

$$\begin{aligned} \overline{\varphi}((a + I) + (b + I)) &= \overline{\varphi}((a + b) + I) \\ &= \varphi(a + b) \\ &= \varphi(a) + \varphi(b) \quad \text{because } \varphi \text{ respects addition} \\ &= \overline{\varphi}(a + I) + \overline{\varphi}(b + I), \end{aligned}$$

so $\overline{\varphi}$ respects addition. Multiplication is proven in exactly the same way.

It remains to prove that $\overline{\varphi}$ is an isomorphism. It is injective, because if $a + I \in \operatorname{Ker}(\overline{\varphi})$, then $0_S = \overline{\varphi}(a + I) = \varphi(a)$, so $a \in \operatorname{Ker}(\varphi) = I$, and so $a + I = I = 0_{R/I}$. Finally, it is surjective because for any $s \in \operatorname{Image}(\varphi)$, there exists $a \in R$ such that $s = \varphi(a) = \overline{\varphi}(a + I)$. □

V − 4 Examples of the First Isomorphism Theorem

We apply Theorem 3.2 to some examples. In each case, we are given a ring homomorphism, and we determine its image and kernel and then apply the theorem. There is no straightforward procedure for finding kernels and images. One has to work with whatever tools are available for the situation at hand. It is useful to remember that the image of a map is equal to its codomain if and only if it is surjective.

Example. Consider the ring homomorphism

$$\begin{aligned} \varphi : \mathbb{Z}[x] &\to \mathbb{Z} \\ f &\mapsto f(0). \end{aligned}$$

It is surjective, because given any $c \in \mathbb{Z}$, the constant polynomial c maps to c. Thus, $\text{Image}(\varphi) = \mathbb{Z}$. As for $\text{Ker}(\varphi)$, we note that $\varphi(f) = 0$ if and only if $f(0) = 0$, if and only if f has zero constant term, if and only if $f = xg$ for some $g \in \mathbb{Z}[x]$. Therefore, $\text{Ker}(\varphi) = x\mathbb{Z}[x]$. By the First Isomorphism Theorem, there is a well-defined isomorphism

$$\begin{aligned} \overline{\varphi} : \mathbb{Z}[x]/x\mathbb{Z}[x] &\to \mathbb{Z} \\ f + x\mathbb{Z}[x] &\mapsto f(0). \end{aligned}$$

Example. Define

$$\begin{aligned} \varphi : \mathbb{Q}[x] &\to \mathbb{Q}^2 \\ f &\mapsto (f(1), f(-1)), \end{aligned}$$

a ring homomorphism. Again, φ is surjective, i.e., $\text{Image}(\varphi) = \mathbb{Q}^2$. To see why, take any $(a, b) \in \mathbb{Q}^2$, and let

$$f = \frac{a}{2}(x + 1) - \frac{b}{2}(x - 1).$$

Then $\varphi(f) = (a, b)$.

We now find $\text{Ker}(\varphi)$. Note that a given $f \in \mathbb{Q}[x]$ is in $\text{Ker}(\varphi)$ if and only $f(1) = f(-1) = 0$. It is clear that the polynomial $(x - 1)(x + 1) = x^2 - 1$ is in $\text{Ker}(\varphi)$, so $(x^2 - 1)_{\mathbb{Q}[x]} \subseteq \text{Ker}(\varphi)$. We claim that in fact $(x^2 - 1)_{\mathbb{Q}[x]} = \text{Ker}(\varphi)$. We will see an easy way to show this once we study polynomial rings in more detail later, but for now, we argue as follows. Suppose that $\text{Ker}(\varphi)$ contained some polynomial not in $(x^2 - 1)_{\mathbb{Q}[x]}$. Then among all such polynomials, there would be one with smallest degree, f say. Note that $\deg(f) \geq 2$ because $f(1) = f(-1) = 0$ and $f \neq 0$, so let $n = \deg(f) - 2 \geq 0$, let c be the leading coefficient of f, and let $g = f - c(x-1)(x+1)x^n$. Then $\deg(g) < \deg(f)$. Now, $g(1) = g(-1) = 0$, so $g \in \text{Ker}(\varphi)$. Also, g is not in $(x^2 - 1)_{\mathbb{Q}[x]}$, for it it were, then $g + c(x^2 - 1)x^n = f$ would be as well, which is not the case. In summary,

- g is in $\text{Ker}(\varphi)$ but not in $(x^2 - 1)_{\mathbb{Q}[x]}$, and
- $\deg(g) < \deg(f)$.

But this contradicts the minimality of the degree of f. Thus, $(x^2-1)_{\mathbb{Q}[x]} = \mathrm{Ker}(\varphi)$, so the First Isomorphism Theorem tells us that $\mathbb{Q}[x]/(x^2-1)_{\mathbb{Q}[x]} \cong \mathbb{Q}^2$.

Example. Let $R = \{a + b\sqrt{2} \in \mathbb{R} \mid a, b \in \mathbb{Z}\}$, a subring of $\mathbb{R}$, and define

$$\begin{aligned} \varphi : R &\to \mathbb{Z}/7\mathbb{Z} \\ a + b\sqrt{2} &\mapsto [a + 3b], \end{aligned}$$

a ring homomorphism. (Note that this map is well defined, for if $a + b\sqrt{2} = a' + b'\sqrt{2}$, where $a, b, a', b' \in \mathbb{Z}$, then $a = a'$ and $b = b'$.) This map is surjective, because for any residue class $[c] \in \mathbb{Z}/7\mathbb{Z}$, we have $\varphi(c) = [c]$. We now find its kernel. Observe that

$$\begin{aligned} \varphi(a + b\sqrt{2}) = [0] &\iff [a + 3b] = [0], \\ &\iff 7 \mid a + 3b, \\ &\iff a = 7c - 3b \text{ for some } c \in \mathbb{Z}. \end{aligned}$$

Thus, $\mathrm{Ker}(\varphi)$ consists of the elements of the form

$$\begin{aligned} & (7c - 3b) + b\sqrt{2} \quad \text{with } b, c \in \mathbb{Z} \\ &= 7c - b(3 - \sqrt{2}) \\ &= (3 + \sqrt{2})(3 - \sqrt{2})c - b(3 - \sqrt{2}) \quad \text{because } 7 = (3 + \sqrt{2})(3 - \sqrt{2}) \\ &= ((3 + \sqrt{2})c - b)(3 - \sqrt{2}), \end{aligned}$$

so $\mathrm{Ker}(\varphi) = \{\alpha(3 - \sqrt{2}) \mid \alpha \in R\} = (3 - \sqrt{2})R$. Therefore, by the First Isomorphism Theorem, there is a well-defined isomorphism

$$\begin{aligned} \overline{\varphi} : R/I &\to \mathbb{Z}/7\mathbb{Z} \\ (a + b\sqrt{2}) + I &\mapsto [a + 3b] \end{aligned}$$

where $I = (3 - \sqrt{2})R$.

Example. Continuing with the previous ring R and ideal I, show that the element $\alpha = (5 - \sqrt{2}) + I$ of R/I is a unit, and find its inverse.

Solution: It will be easier to work in the ring $\mathbb{Z}/7\mathbb{Z}$, which we know from the previous example is isomorphic to R/I via $\overline{\varphi}$. We have

$$\overline{\varphi}(\alpha) = \varphi(5 - \sqrt{2}) = [5 - 3] = [2],$$

which is invertible in $\mathbb{Z}/7\mathbb{Z}$ with inverse $[4]$. That is, $[4][2] = [1]$. Therefore, applying $\overline{\varphi}^{-1}$ to both sides, we obtain

$$\begin{aligned} & \overline{\varphi}^{-1}([4])\overline{\varphi}^{-1}([2]) = \overline{\varphi}^{-1}([1]), \\ & \text{i.e.,} \quad (4 + I)\alpha = 1 + I. \end{aligned} \tag{4.1}$$

Therefore, $\alpha^{-1} = 4 + I$.

Exercise. Verify (4.1) directly by computing $(4+I)\alpha$ in the ring R/I and showing that it is equal to $1 + I$. Remember that $(3 + \sqrt{2})(3 - \sqrt{2}) = 7$.

(VI) Divisibility and Factorization

VI – 1 Introduction

Historically, the notion of divisibility began with the integers. But as mathematics has developed, divisibility has been found to play an important role in a variety of contexts, not least in polynomial rings and in rings of *algebraic integers*, such as $\{x + y\sqrt{2} \in \mathbb{R} \mid x, y \in \mathbb{Z}\}$.

If a and b are elements in a commutative ring R, then b is said to *divide* a if there is $c \in R$ such that $a = bc$, just as in $\mathbb{Z}$, and we write $b \,|\, a$ in this case. For example, in $\mathbb{Z}[x]$, $x + 2$ divides $x^2 - x - 6$, because $x^2 - x - 6 = (x + 2)(x - 3)$.

As soon as divisibility is at work, a natural consideration is the problem of factorization: whether it is possible to factorize a ring element—in some suitable way—and, if so, whether a factorization is unique in an appropriate sense.

Again, the integers are the model here. We know that positive integers can be factorized into products of prime numbers, and that such a factorization is essentially unique, in the way described in Theorem 3.2 in Section II, the Fundamental Theorem of Arithmetic. We will give formal definitions of factorization and uniqueness of factorization in due course, but for now imagine something along the lines of the properties described in that theorem.

To this day, much mystery surrounds the question of which rings have uniqueness of factorization. For example, the ring $R = \{x + y\sqrt{2} \in \mathbb{R} \mid x, y \in \mathbb{Z}\}$ does, while the ring $S = \{x + y\sqrt{10} \in \mathbb{R} \mid x, y \in \mathbb{Z}\}$ does not, yet the definitions of the rings R and S differ only in that one involves the number 2 while the other involves 10. What property does 2 have, and that 10 does not, to adequately explain the difference? The answer is not well understood at any deep level. Even among the rings in a certain natural family to which R and S belong, we know of no simple description of the ones that do have uniqueness of factorization and the ones that do not.

In this final section of the course, we build up some methodology to prove uniqueness of factorization at least in some special classes of ring. The idea is roughly as follows:

(i) If R is a *principal ideal domain* (the essential property being that all of its ideals are principal), then R has uniqueness of factorization.

(ii) If R is a *Euclidean domain* (the essential property being that the Euclidean algorithm can be performed in it), then it is a principal ideal domain and therefore has uniqueness of factorization.

VI – 2 Integral domains and fields

A convenient setting in which to tackle problems of divisibility and factorization is that of an *integral domain*. An integral domain is a commutative ring R such that

(i) R has an identity $1 \neq 0$, and

(ii) for all $a, b \in R$ such that $ab = 0$, either $a = 0$ or $b = 0$.

Remark. Property (ii) is equivalent to saying that, for all $a, b \in R \smallsetminus \{0\}$, $ab \neq 0$. That is, a product of non-zero elements is non-zero.

Example. The rings $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ are all integral domains.

Example. The polynomial rings $\mathbb{Z}[x]$, $\mathbb{Q}[x]$, and $\mathbb{R}[x]$ are all integral domains. More generally, if R is an integral domain, then so is $R[x]$. Let us prove this. Suppose that $f = a_m x^m + \cdots + a_0$ and $g = b_n x^n + \cdots + b_0$ are elements of $R[x]$, where $a_m, b_n \neq 0$. Then the coefficient of x^{m+n} in fg is $a_m b_n$, and this is non-zero because R is an integral domain.

Example. The ring $\mathbb{Z} \times \mathbb{Z}$ is not an integral domain, because $(1, 0)$ and $(0, 1)$ are two non-zero elements whose product is zero. More generally, if R and S are non-zero rings, then $R \times S$ is not an integral domain.

Example. The ring $\mathbb{Z}/6\mathbb{Z}$ is not an integral domain, because $[2][3] = [0]$.

Fields

We have a related notion, that of a *field*. A field is a commutative ring R such that

(i) R has an identity $1 \neq 0$, and

(ii) for all $a \in R \smallsetminus \{0\}$, there is $b \in R$ such that $ba = 1$.

Remark. In the presence of the assumption of commutativity, property (ii) is equivalent to saying that $R^\times = R \smallsetminus \{0\}$, i.e., every non-zero element is a unit.

Example. The ring $\mathbb{Z}/5\mathbb{Z}$ is a field. We will see a more fundamental reason for this in Proposition 2.2 below, but for now, let us verify the assertion simply by computing the multiplication table:

$\cdot$	[0]	[1]	[2]	[3]	[4]
[0]	[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]	[3]	[4]
[2]	[0]	[2]	[4]	[1]	[3]
[3]	[0]	[3]	[1]	[4]	[2]
[4]	[0]	[4]	[3]	[2]	[1]

Every row other than the row of zeroes has a [1] in it, which is to say that every non-zero element of $\mathbb{Z}/5\mathbb{Z}$ has an inverse.

Example. The rings $\mathbb{Q}$ and $\mathbb{R}$ are fields. However, $\mathbb{Z}$ is not a field, because its only units are 1 and -1.

Proposition 2.1. *Every field is an integral domain.*

Proof. Let F be a field. By assumption, F is commutative and has a non-zero identity, so we have only to check the condition that if $ab = 0$, then $a = 0$ or $b = 0$. If $ab = 0$ and $a \neq 0$, then because F is a field, a has an inverse, c say. Then $0 = c(ab) = (ca)b = 1 \cdot b = b$. □

Exercise. It is in fact true that every *finite* integral domain is in turn a field. Prove this.

We saw in examples above that $\mathbb{Z}/5\mathbb{Z}$ is a field, and therefore an integral domain, but that $\mathbb{Z}/6\mathbb{Z}$ is not an integral domain, and therefore not a field. The criterion for deciding when $\mathbb{Z}/n\mathbb{Z}$ is a field is straightforward.

Proposition 2.2. *Let $n \geq 2$ be an integer. Then the following are equivalent:*

(i) n is prime.

(ii) $\mathbb{Z}/n\mathbb{Z}$ is an integral domain.

(iii) $\mathbb{Z}/n\mathbb{Z}$ is a field.

Proof. Recall from Proposition 7.1 in Section IV that

$$(\mathbb{Z}/n\mathbb{Z})^\times = \{[a] \in \mathbb{Z}/n\mathbb{Z} \mid \gcd(a, n) = 1\}.$$

If n is prime, then every integer in $\{1, \ldots, n-1\}$ is coprime to n, so $(\mathbb{Z}/n\mathbb{Z})^\times = \mathbb{Z}/n\mathbb{Z} \smallsetminus \{[0]\}$, and so $\mathbb{Z}/n\mathbb{Z}$ is a field. This shows that (i) $\Rightarrow$ (iii). Next, if $\mathbb{Z}/n\mathbb{Z}$ is a field, then it is an integral domain by Proposition 2.1, so (iii) $\Rightarrow$ (ii). Finally, assume that $\mathbb{Z}/n\mathbb{Z}$ is an integral domain. We show that n is prime. If $n = ab$, where $a, b \in \mathbb{Z}_{>0}$, then in $\mathbb{Z}/n\mathbb{Z}$,

$$[0] = [n] = [ab] = [a][b],$$

so because $\mathbb{Z}/n\mathbb{Z}$ is an integral domain, either $[a] = [0]$ or $[b] = [0]$. Thus, either $n \mid a$ (in which case $a = n$), or $n \mid b$ (in which case $b = n$). □

Exercise. Show that in a field F, the only ideals are $\{0\}$ and F.

In the context of divisibility, the notions of *integral domain* and *field* are two extremes. The former notion is very broad: If all we know about a ring is that it is an integral domain, there is not much we can say about divisibility in the ring, because integral domains include many flavours of ring, exhibiting collectively a wide range of behaviours with respect to divisibility and factorization. On the other hand, a field is a very particular type of integral domain, in which divisibility is straightforward: Every non-zero element of a field F divides every element of F. For the purposes at hand, the interesting families of rings are the ones that lie in between, and to these we now turn.

VI – 3 Complex numbers and quadratic rings

Now that we have introduced the notions of integral domain and field, we set aside a section to introduce some especially important examples.

Complex numbers

There are several constructions of the complex numbers, but here is a straightforward one. Let $\mathbb{C}$ be the set $\mathbb{R}^2$, that is, the set of pairs (a, b) where $a, b \in \mathbb{R}$, together with the following operations:

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

Note that addition is defined componentwise, as in the product ring $\mathbb{R} \times \mathbb{R}$, but that multiplication is different. We leave it as an exercise to show that $\mathbb{C}$ is a ring with respect to these operations. Its elements are called *complex numbers*.

The ring $\mathbb{C}$ is commutative and unital, the identity element being $(1, 0)$. By a slight abuse of notation, if $a \in \mathbb{R}$, then the element $(a, 0)$ of $\mathbb{C}$ will be abbreviated to simply a. Further, if we let $i = (0, 1)$, then for all $b \in \mathbb{R}$,

$$\begin{aligned}(0, b) &= (b, 0)(0, 1) \quad \text{by the multiplication law} \\ &= bi \quad \text{because of our convention that } b \text{ represents the element } (b, 0).\end{aligned}$$

Thus, every element of $\mathbb{C}$ may be expressed uniquely in the form $a + bi$ where $a, b \in \mathbb{R}$. In this notation, addition and multiplication are as follows:

$$(a + bi) + (c + di) = (a + c) + (b + d)i,$$
$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i.$$

The real numbers a and b in a complex number $\alpha = a + bi$ are called, respectively, the *real part* and the *imaginary part* of α, denoted $\mathrm{Re}(\alpha)$ and $\mathrm{Im}(\alpha)$. Thus, for any $\alpha \in \mathbb{C}$, $\alpha = \mathrm{Re}(\alpha) + \mathrm{Im}(\alpha)i$.

We may show that $\mathbb{C}$ is a field by observing that, if $a, b \in \mathbb{R}$ are not both zero, then

$$\left(\frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2}i\right)(a + bi) = 1.$$

We call $\mathbb{C}$ the *field of complex numbers*. It contains square roots of -1: $(\pm i)^2 = -1$.

Complex conjugation

If $\alpha = a + bi$, where $a, b \in \mathbb{R}$, then the *complex conjugate* of α is the complex number $\overline{\alpha} = a - bi$. Complex conjugation is a ring homomorphism: $\overline{\alpha + \beta} = \overline{\alpha} + \overline{\beta}$, and $\overline{\alpha\beta} = \overline{\alpha}\,\overline{\beta}$ for all $\alpha, \beta \in \mathbb{C}$.

Observe that $\alpha\,\overline{\alpha} = a^2 + b^2$, a non-negative real number, and its square root is called the *modulus* or *absolute value* of α, denoted $|\alpha|$. Thus,

$$|\alpha| = \sqrt{\alpha\,\overline{\alpha}} = \sqrt{a^2 + b^2}.$$

Because complex conjugation respects multiplication, it follows that the modulus does as well: $|\alpha\beta| = |\alpha|\,|\beta|$.

One of the most important properties of the field of complex numbers is expressed in the following.

Theorem 3.1 (Fundamental Theorem of Algebra)**.** *Every non-constant polynomial in* $\mathbb{C}[x]$ *has a root in* $\mathbb{C}$*.*

The proof is beyond the scope of this course. In any case, we will not need the theorem. All we will need for our purposes is the existence of roots of quadratic polynomials with real coefficients. If $x^2 + bx + c \in \mathbb{R}[x]$, then its roots are $\frac{1}{2}(-b \pm \sqrt{b^2 - 4c})$ if $b^2 - 4c \geq 0$, and $\frac{1}{2}(-b \pm \sqrt{4c - b^2}\,i)$ if $b^2 - 4c < 0$.

Remark. In algebra, it is often immaterial which particular square root the symbol $\sqrt{a}$ denotes. However, for concreteness, let us make a choice. When a is a non-negative real number, $\sqrt{a}$ will mean the non-negative real square root. When a is instead negative, $\sqrt{a}$ will mean the complex number $\sqrt{-a}\,i$. (Note that $(\sqrt{-a}\,i)^2 = -a(-1) = a$.)

Quadratic rings

Consider the ring $\{x + y\sqrt{2} \in \mathbb{R} \mid x, y \in \mathbb{Z}\}$, an integral domain. We will denote it $\mathbb{Z}[\sqrt{2}]$. Because $\sqrt{2}$ is not rational, it follows that every element of $\mathbb{Z}[\sqrt{2}]$ can be written *uniquely* in the form $x + y\sqrt{2}$ with $x, y \in \mathbb{Z}$. This, and other rings like it, will be an important source of examples. We generalize $\mathbb{Z}[\sqrt{2}]$ as follows.

Fix an integer n that is not a square, i.e., such that the equation $x^2 - n = 0$ has no integer solution x. Examples of such integers n are $2, 3, 5, 6, 7, 8, 10$, and also any negative integer. Because n is not square, $\sqrt{n}$ is not a rational number: it is an irrational real number when $n > 0$, and a non-real complex number when $n < 0$.

We define $\mathbb{Z}[\sqrt{n}] = \{x + y\sqrt{n} \mid x, y \in \mathbb{Z}\}$, an integral domain. It is a subring of $\mathbb{R}$ when $n > 0$, and a subring of $\mathbb{C}$ no matter what n is. Again, because n is not square, every element of $\mathbb{Z}[\sqrt{n}]$ can be written *uniquely* in the form $x + y\sqrt{n}$ with $x, y \in \mathbb{Z}$.

The importance of this ring for number theory lies in the fact that an equation of the form $x^2 - ny^2 = k$, where k is some fixed integer, can be written instead as $(x - y\sqrt{n})(x + y\sqrt{n}) = k$, the left-hand side being now a product in the ring $\mathbb{Z}[\sqrt{n}]$.

A useful map on $\mathbb{Z}[\sqrt{n}]$ is the *norm*, defined by

$$\begin{aligned} N : \mathbb{Z}[\sqrt{n}] &\to \mathbb{Z} \\ x + y\sqrt{n} &\mapsto x^2 - ny^2 = (x + y\sqrt{n})(x - y\sqrt{n}). \end{aligned}$$

Exercise.

(i) Show that N respects multiplication: If $\alpha, \beta \in \mathbb{Z}[\sqrt{n}]$, then $N(\alpha\beta) = N(\alpha)N(\beta)$.

(ii) Deduce that $\alpha \in \mathbb{Z}[\sqrt{n}]^\times$ if and only if $N(\alpha) \in \{1, -1\}$.

VI – 4 Euclidean domains

Let R be an integral domain. A *Euclidean function* on R is a map $\phi : R \to \mathbb{Z}_{\geq 0}$ such that

(i) $\phi(0) = 0$, and

(ii) for all $a, b \in R$ with $b \neq 0$, there are $q, r \in R$ such that $a = qb + r$ and either $r = 0$ or $\phi(r) < \phi(b)$.

A *Euclidean domain* is an integral domain that has a Euclidean function. Thus, in a Euclidean domain, one may perform "division with remainder" where the remainder either is zero or is smaller, in some sense, than the element one is dividing by.

Example. The ring $\mathbb{Z}$ is a Euclidean domain, because the function $\phi : a \mapsto |a|$ is a Euclidean function on $\mathbb{Z}$ by Proposition 1.1 in Section II. Specifically, we proved that if $a, b \in \mathbb{Z}$ and $b \neq 0$, then there are $q, r \in \mathbb{Z}$ with $0 \leq r < |b|$ such that $a = qb + r$. (We even showed that q and r are unique, but this need not be the case in other Euclidean domains.)

Example. If F is a field, then $F[x]$ is a Euclidean domain with Euclidean function

$$\phi(f) = \begin{cases} \deg(f) & \text{if } f \neq 0, \\ 0 & \text{if } f = 0. \end{cases}$$

One may prove by induction that this ϕ is a Euclidean function. However, rather than doing so here, we will illustrate the idea via *long division of polynomials*. If, for example, $f = 3x^5 + 4x^4 - x^3 + 2x^2 + 3x + 7$ and $g = x^2 - 2x + 2$, polynomials in $\mathbb{Q}[x]$, then the quotient q and remainder r in the expression $f = qg + r$ are found as follows:

$$\begin{array}{rr|l}
 & & 3x^3 + 10x^2 + 13x + 8 \\
\hline
x^2 - 2x + 2 &) & 3x^5 + 4x^4 - x^3 + 2x^2 + 3x + 7 \\
 & & -3x^5 + 6x^4 - 6x^3 \\
\hline
 & & 10x^4 - 7x^3 + 2x^2 \\
 & & -10x^4 + 20x^3 - 20x^2 \\
\hline
 & & 13x^3 - 18x^2 + 3x \\
 & & -13x^3 + 26x^2 - 26x \\
\hline
 & & 8x^2 - 23x + 7 \\
 & & -8x^2 + 16x - 16 \\
\hline
 & & -7x - 9
\end{array}$$

That is, $q = 3x^3 + 10x^2 + 13x + 8$ and $r = -7x - 9$.

We will discuss the above algorithm in more detail in class. Note that, as in the case of $\mathbb{Z}$, the quotient q and remainder r in long division of polynomials are unique, unlike in some other Euclidean domains.

Example. Let us do another example, again in $\mathbb{Q}[x]$. Here, we find the quotient q and remainder r on dividing $f = 3x^4 - 5x^3 - x^2 + x + 2$ by $g = 2x^2 - 3x + 5$:

$$\begin{array}{rl} & \qquad\qquad \tfrac{3}{2}x^2 \;-\tfrac{1}{4}x \;-\tfrac{37}{8} \\ \hline 2x^2 - 3x + 5\big) & \; 3x^4 - 5x^3 \;\;- x^2 \;\;\;+ x \;\;+ 2 \\ & \underline{-3x^4 + \tfrac{9}{2}x^3 - \tfrac{15}{2}x^2} \\ & \qquad -\tfrac{1}{2}x^3 - \tfrac{17}{2}x^2 \;\;+ x \\ & \qquad \underline{\tfrac{1}{2}x^3 \;-\tfrac{3}{4}x^2 \;+\tfrac{5}{4}x} \\ & \qquad\qquad -\tfrac{37}{4}x^2 \;+\tfrac{9}{4}x \;\;+ 2 \\ & \qquad\qquad \underline{\tfrac{37}{4}x^2 - \tfrac{111}{8}x + \tfrac{185}{8}} \\ & \qquad\qquad\qquad -\tfrac{93}{8}x + \tfrac{201}{8} \end{array}$$

Thus, $f = qg + r$ where $q = \frac{3}{2}x^2 - \frac{1}{4}x - \frac{37}{8}$ and $r = -\frac{93}{8}x + \frac{201}{8}$.

Example. If $x, y \in \mathbb{Q}$, define $\phi(x + y\sqrt{2}) = |x^2 - 2y^2|$. We show that ϕ restricts to a Euclidean function $\mathbb{Z}[\sqrt{2}] \to \mathbb{Z}_{\geq 0}$, so that $\mathbb{Z}[\sqrt{2}]$ is a Euclidean domain. We will use the fact that $\phi((x + y\sqrt{2})(z + w\sqrt{2})) = \phi(x + y\sqrt{2})\phi(z + w\sqrt{2})$ for all $x, y, z, w \in \mathbb{Q}$.

Suppose that $\alpha, \beta \in \mathbb{Z}[\sqrt{2}]$ and $\beta \neq 0$. Then by rationalizing the denominator in α/β, we may write

$$\alpha/\beta = x + y\sqrt{2} \quad \text{where } x, y \in \mathbb{Q}. \tag{4.1}$$

Choose $x', y' \in \mathbb{Z}$ such that $|x - x'| \leq 1/2$ and $|y - y'| \leq 1/2$, and let $\gamma = x' + y'\sqrt{2} \in \mathbb{Z}[\sqrt{2}]$. Then

$$\begin{aligned} \alpha - \gamma\beta &= (x + y\sqrt{2})\beta - \gamma\beta \quad \text{by (4.1)} \\ &= (x + y\sqrt{2})\beta - (x' + y'\sqrt{2})\beta = ((x - x') + (y - y')\sqrt{2})\beta. \end{aligned}$$

Hence,

$$\begin{aligned} \phi(\alpha - \gamma\beta) &= \phi((x - x') + (y - y')\sqrt{2})\phi(\beta) \quad \text{because } \phi \text{ respects multiplication} \\ &= |(x - x')^2 - 2(y - y')^2|\phi(\beta) \\ &\leq \max((x - x')^2, 2(y - y')^2)\phi(\beta) \leq \frac{1}{2}\phi(\beta) < \phi(\beta). \end{aligned}$$

Thus, if $\rho = \alpha - \gamma\beta$, then $\alpha = \gamma\beta + \rho$ with $\phi(\rho) < \phi(\beta)$.

For example, if $\alpha = 1 + 3\sqrt{2}$ and $\beta = 4 - \sqrt{2}$, then

$$\frac{\alpha}{\beta} = \frac{1 + 3\sqrt{2}}{4 - \sqrt{2}} = \frac{(1 + 3\sqrt{2})(4 + \sqrt{2})}{(4 - \sqrt{2})(4 + \sqrt{2})} = \frac{5}{7} + \frac{13}{14}\sqrt{2}.$$

The closest integers to $5/7$ and $13/14$ are 1 and 1, so we let $\gamma = 1 + \sqrt{2}$ and

$$\rho = \alpha - \gamma\beta = (1 + 3\sqrt{2}) - (1 + \sqrt{2})(4 - \sqrt{2}) = -1.$$

Then $\alpha = \gamma\beta + \rho$ with $\phi(\rho) < \phi(\beta)$.

VI – 5 Principal ideal domains

A *principal ideal domain* (or *PID*) is an integral domain in which every ideal is principal.

We saw in Section IV – 9 that the ideal $(2, x)_{\mathbb{Z}[x]}$ of $\mathbb{Z}[x]$ is not principal, so not every integral domain is a principal ideal domain. However, principal ideal domains do exist, as the next proposition shows.

Proposition 5.1. *Every Euclidean domain is a principal ideal domain.*

Proof. Let R be a Euclidean domain and $\phi : R \to \mathbb{Z}_{\geq 0}$ a Euclidean function on R. Let I be an ideal of R. If $I = \{0\}$, then I is principal, because $\{0\} = (0)_R$. We may assume, then, that I contains non-zero elements. Among all the non-zero elements a of I, choose one such that $\phi(a)$ is least, and call it m.

Now, given any $a \in I$, write it as $a = qm + r$ where $q, r \in R$ and either $r = 0$ or $\phi(r) < \phi(m)$. Because $a, m \in I$, and because I is closed under subtraction and under multiplication by ring elements, it follows that $r = a - qm \in I$. Therefore, if r were non-zero, then by the minimality of $\phi(m)$, we would have to have $\phi(r) \geq \phi(m)$, contradicting the remainder property of r. Consequently, $r = 0$, so $a = qm \in I$. Thus, $I = (m)_R$. □

In light of the examples we saw of Euclidean domains, we know by Proposition 5.1 that all of the following are principal ideal domains:

- $\mathbb{Z}$
- $F[x]$ where F is a field
- $\mathbb{Z}[\sqrt{2}]$

A field F is also a principal ideal domain, because its only ideals are $\{0\} = (0)_F$ and $F = (1)_F$. Another, but far less direct, way to see that F is a principal ideal domain is to show that it is a Euclidean domain (exercise) and use Proposition 5.1.

Remark. Because $\mathbb{Q}[x]$ is a principal ideal domain but $\mathbb{Z}[x]$ is not, we see that a subring of a PID need not be a PID.

The hierarchy so far

We have so far seen that

- every field is a Euclidean domain (this was left as an exercise, but it is a short one),
- every Euclidean domain is a principal ideal domain, and
- every principal ideal domain is an integral domain (simply by definition).

To continue this story, we need to unpack divisibility a little further.

VI – 6 Prime and irreducible elements

Let R be an integral domain. Recall that the notation $b \mid a$ means "b divides a".

(i) A *prime element* is an element $r \in R$ satisfying all of the following: (i) $r \neq 0$, (ii) $r \notin R^\times$, and (iii) for all $a, b \in R$ such that $r \mid ab$, either $r \mid a$ or $r \mid b$.

(ii) An *irreducible element* is an element $r \in R$ satisfying all of the following: (i) $r \neq 0$, (ii) $r \notin R^\times$, and (iii) for all $a, b \in R$ such that $r = ab$, either $a \in R^\times$ or $b \in R^\times$.

Note how the definition of prime here is like the property of prime numbers that we proved in Lemma 3.1. The difference now is that we take this property as a definition.

It is also worth observing that the definition of an irreducible element resembles the familiar definition of a prime number, if we remember that $\mathbb{Z}^\times = \{1, -1\}$. If this causes confusion, just remember that

- the notions of *prime element* and *irreducible element* in a general integral domain are defined by (i) and (ii) above, and
- in the ring $\mathbb{Z}$, the two notions coincide and correspond to the numbers $\pm p$ where p is a prime number in the usual sense.

Example. In $\mathbb{Z}[x]$, x is prime. Indeed, suppose that $f, g \in \mathbb{Z}[x]$ are polynomials such that $x \mid fg$, i.e., $fg = xh$ for some polynomial $h \in \mathbb{Z}[x]$. Then

$$f(0)g(0) = (fg)(0) = 0 \cdot h(0) = 0,$$

so either $f(0) = 0$ or $g(0) = 0$, and so either $x \mid f$ or $x \mid g$.

In fact, x is also an irreducible element of $\mathbb{Z}[x]$. This follows from Proposition 6.1 below, but we may see it directly as follows. Suppose $x = fg$, where $f, g \in \mathbb{Z}[x]$. Then $\deg(f) + \deg(g) = \deg(x) = 1$, so either $\deg(f) = 0$ or $\deg(g) = 0$. If $\deg(f) = 0$, then f is equal to a non-zero constant, a say, but then if b is the coefficient of x in g, we have $ab = 1$, so $a \in \{1, -1\} = \mathbb{Z}[x]^\times$. Similarly, if $\deg(g) = 0$, then g is a unit in $\mathbb{Z}[x]$.

Example. Show that 3 is irreducible in $\mathbb{Z}[\sqrt{-1}] = \mathbb{Z}[i]$.

Solution: Suppose that $3 = \alpha\beta$, where $\alpha, \beta \in \mathbb{Z}[i]$. Taking norms, we obtain $9 = N(\alpha\beta) = N(\alpha)N(\beta)$, so $N(\alpha) \mid 9$. If $N(\alpha) = \pm 1$, then α is a unit, and if $N(\beta) = \pm 1$, then β is. It remains to rule out the possibility that $N(\alpha) = \pm 3$. But if $\alpha = x + yi$ with $x, y \in \mathbb{Z}$, then $N(\alpha) = \pm 3$ says $x^2 + y^2 = \pm 3$, which has no integer solutions. Thus, 3 is irreducible.

In the proof of the next proposition, we will use the following cancellation property for integral domains: If a, b, c are elements in an integral domain R such that $ab = ac$, and if $a \neq 0$, then $b = c$. We leave this as a short exercise. It uses only the definition of an integral domain.

Proposition 6.1. *Let R be an integral domain. Then every prime element of R is irreducible.*

Proof. Let r be a prime element of R, and suppose that $r = ab$ where $a, b \in R$. Then $r \mid ab$ (because $ab = r \cdot 1$), so because r is prime, r divides a or b. If r divides a, which is to say that $a = rd$ for some $d \in R$, then $r = ab = rdb$. Hence, remembering that $r \neq 0$, and remembering also that R is an integral domain, we may cancel r from both sides to obtain $1 = db$, so $b \in R^\times$. Similarly, if r divides b, then the same argument shows that $a \in R^\times$. □

Irreducible elements in PIDs

We know that a prime element in an integral domain is always irreducible. What about the converse? Is an irreducible element always prime? In fact, the answer is no. In the ring $\{x + y\sqrt{10} \in \mathbb{R} \mid x, y \in \mathbb{Z}\}$, for example, the element $7 + \sqrt{10}$ is irreducible but not prime. I will give some hints in class for how to show this.

However, in a principal ideal domain, irreducible elements are always prime, as we will see after the following lemma.

Lemma 6.2. *If R is a commutative ring with identity $1 \neq 0$, and if $a \in R$, then a is a unit in R if and only if $(a)_R = R$.*

Proof. If $(a)_R = R$, then in particular $1 \in (a)_R = Ra$, so there is $b \in R$ such that $1 = ba$, i.e., $a \in R^\times$. Conversely, if $a \in R^\times$ and b is any element of R, then $b = b \cdot 1 = ba^{-1}a \in Ra = (a)_R$, so $R = (a)_R$. □

Proposition 6.3. *Let R be a principal ideal domain. Then every irreducible element of R is prime. (Thus, in a PID, the notions of* prime element *and* irreducible element *are interchangeable.)*

Proof. Let π be an irreducible element of R, and suppose that $\pi \mid ab$ where $a, b \in R$. We show that either $\pi \mid a$ or $\pi \mid b$. Since R is a PID, there is $d \in R$ such that $(\pi, a)_R = (d)_R$. Therefore, because $\pi \in (\pi, a)_R$, $\pi = d\pi'$ for some $\pi' \in R$, and similarly, $a = da'$ for some $a' \in R$.

Now, π is irreducible, so either $d \in R^\times$ or $\pi' \in R^\times$. If the latter, then $a = da' = \pi(\pi')^{-1}a'$, and so $\pi \mid a$. If the former, i.e., if $d \in R^\times$, then $(d)_R = R$ by Lemma 6.2, i.e., $(\pi, a)_R = R$. In particular, $1 \in (\pi, a)_R$, so there are $m, n \in R$ such that $1 = m\pi + na$. Hence, $b = m\pi b + nab$, which is divisible by π because $\pi \mid ab$ by assumption. □

Example. Let $t \in \mathbb{Q}$. Show that the polynomial $x - t \in \mathbb{Q}[x]$ is prime.

Solution: Because $\mathbb{Q}[x]$ is a principal ideal domain (see Section 5), it is enough to show that $x - t$ is irreducible. Suppose that $x - t = fg$, where $f, g \in \mathbb{Q}[x]$. Then $\deg(f) + \deg(g) = \deg(x - t) = 1$, so either $\deg(f) = 0$ or $\deg(g) = 0$. If $\deg(f) = 0$, then f is equal to some non-zero constant a and is therefore a unit in $\mathbb{Q}[x]$. Similarly, if $\deg(g) = 0$, then g is a unit in $\mathbb{Q}[x]$.

Example. The ring $\mathbb{Z}[i]$ is a Euclidean domain (mimic the last example in Section 4), so it is a principal ideal domain. We saw just above that 3 is irreducible in $\mathbb{Z}[i]$, so it is also prime in $\mathbb{Z}[i]$.

VI – 7 Irreducibility in polynomial rings

When F is a field, there are several straightforward techniques to help decide whether a given polynomial in $F[x]$ is irreducible.

Proposition 7.1. *Let F be a field, $f \in F[x]$, and $a \in F$. Then $f(a) = 0$ if and only if $x - a$ divides f in $F[x]$.*

Proof. If $x - a$ divides f, i.e., $f = (x - a)g$ for some $g \in F[x]$, then $f(0) = 0 \cdot g(0) = 0$. Conversely, suppose that $f(a) = 0$. Applying division with remainder, we may write $f = (x - a)q + r$ where $q, r \in F[x]$ and $\deg(r) < \deg(x - a) = 1$. Thus, r is an element of F, so evaluating both sides of the equation $f = (x - a)q + r$ at $x = a$ gives $0 = r$. Thus, $f = (x - a)q$. □

Example. Show that the polynomial $f = x^4 - x^3 - 3x^2 + 5x - 6 \in \mathbb{Q}[x]$ is not irreducible.

Solution: We see that $f(2) = 0$, so $f = (x-2)g$ for some $g \in \mathbb{Q}[x]$. Further, neither $x-2$ nor g is a unit in $\mathbb{Q}[x]$. Indeed, the units in $\mathbb{Q}[x]$ are the non-zero constant polynomials, i.e., the polynomials of degree 0, but $x - 2$ has degree 1 and g has degree 3.

How did we know to try evaluating f at 2 in the previous example, among all the infinite possible numbers to evaluate at? If a non-zero polynomial has integer coefficients, then one may determine the possibilities for, at least, the rational roots by looking at the leading coefficient and the constant term:

Proposition 7.2. *Let $f = a_n x^n + \cdots + a_1 x + a_0 \in \mathbb{Z}[x]$, where $n \geq 1$ and $a_n \neq 0$. If r/s is a root of f, where r and s are coprime integers and $s \neq 0$, then $r \mid a_0$ and $s \mid a_n$.*

Proof. If we multiply both sides of the equation $f(r/s) = 0$ by s^n, we arrive at

$$a_n r^n + a_{n-1} r^{n-1} s + \cdots + a_1 r s^{n-1} + a_0 s^n = 0.$$

Therefore, $r \mid a_0 s^n$ and $s \mid a_n r^n$. Hence, because r and s are coprime, part (i) of the exercise in Section II – 3 shows that $r \mid a_0$ and $s \mid a_n$. □

Example. The polynomial $2x^{23} + 17x^{11} - 80x^4 - 3$ has no rational roots. Indeed, by Proposition 7.2, we have only to try the eight potential roots $\pm 1, \pm 3, \pm 1/2, \pm 3/2$, and none of these is actually a root.

Proposition 7.3. *Let F be a field, and suppose that $f \in F[x]$ has degree 2 or 3. Then f is irreducible in $F[x]$ if and only if f has no root in F.*

Proof. If f has a root $a \in F$, then $f = (x - a)g$ for some $g \in F[x]$ by Proposition 7.1. Because $\deg(x - a)$ and $\deg(g)$ are both positive, neither $x - a$ nor g is a unit in $F[x]$, so f is not irreducible.

Conversely, suppose that f is not irreducible. By assumption, f has degree 2 or 3, so in particular it is non-zero and not a unit, so the assumption that it is not irreducible implies that $f = gh$ for some $g, h \in F[x]$ where g and h are not units (and not zero).

Therefore, $\deg(g)$ and $\deg(h)$ are both at least 1, but we also have $\deg(g) + \deg(h) = \deg(f) \in \{2, 3\}$, so at least one of g, h has degree 1, i.e., is a linear factor. But a linear polynomial has a root in F, so f does as well. □

Example. Show that the polynomial $f = x^3 + 3x^2 + 3x - 1 \in \mathbb{Q}[x]$ is irreducible.

Solution: We show that f has no rational roots, so that we may apply Proposition 7.3 to it to deduce that it is irreducible. If f had a rational root, then by Proposition 7.2, it would have to be 1 or -1. However, $f(1) = 6 \neq 0$, and $f(-1) = -2 \neq 0$, so f has no roots in $\mathbb{Q}$.

Caution. If $f \in F[x]$ has degree 4 or more, then Proposition 7.3 does not apply. For example, the polynomial $f = x^4 - 4 \in \mathbb{Q}[x]$ has no roots in $\mathbb{Q}$, as we may see via Proposition 7.2, but neither is it irreducible, for $f = (x^2 - 2)(x^2 + 2)$. Nonetheless, it is always true that if $f \in F[x]$ has degree at least 2 and has a root in F, then it is not irreducible.

A *monic* polynomial is a non-zero polynomial whose leading coefficient is 1.

Theorem 7.4 (Eisenstein's Criterion)**.** *Let $f = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in \mathbb{Z}[x]$ be a monic polynomial of degree at least 1, and suppose that there is a prime number p such that $p \mid a_i$ for all $i \in \{0, \ldots, n-1\}$ but $p^2 \nmid a_0$. Then f is irreducible in both $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$.*

For a proof, see Section 4 of the Appendix. A polynomial meeting the assumptions of Theorem 7.4 is called an *Eisenstein polynomial*. The theorem says that all Eisenstein polynomials are irreducible.

Example. By Eisenstein's Criterion with the prime $p = 3$, the polynomial $f = x^5 - 6x^4 + 21x^3 + 9x^2 + 162x + 105$ is irreducible in both $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$.

Example. By Eisenstein's Criterion with the prime $p = 7$, the polynomial $f = x^{10\,001} + 49x^{310} + 35$ is irreducible in both $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$.

Exercise. Let $f = 2x^2 - 10x + 6 \in \mathbb{Z}[x]$. Decide whether f is irreducible in $\mathbb{Q}[x]$, and decide also whether it is irreducible in $\mathbb{Z}[x]$.

Exercise. Prove that the polynomial $g = x^3 - 3x^2 + 3x + 3^2$ is irreducible in $\mathbb{Q}[x]$. This shows that an irreducible polynomial in $\mathbb{Q}[x]$ need not be Eisenstein.

VI – 8 Prime ideals

Let R be a commutative ring with identity $1 \neq 0$. An ideal P of R is called *prime* if it safisfies both of the following: (i) $P \neq R$, and (ii) for all $a, b \in R$ such that $ab \in P$, either $a \in P$ or $b \in P$.

Example. The ideal $(5)_{\mathbb{Z}} = 5\mathbb{Z}$ in $\mathbb{Z}$ is prime. Indeed, if $ab \in 5\mathbb{Z}$, then $5 \,|\, ab$, so either $5 \,|\, a$ or $5 \,|\, b$, because 5 is a prime number. (Use Lemma 3.1 in Section II.) Thus, $a \in 5\mathbb{Z}$ or $b \in 5\mathbb{Z}$.

Example. The ideal $(6)_{\mathbb{Z}}$ is not a prime ideal of $\mathbb{Z}$. For example, $2 \cdot 3 = 6 \in 6\mathbb{Z}$, but neither 2 nor 3 is in $6\mathbb{Z}$.

Exercise. Let R be an integral domain, and let $a \in R \setminus \{0\}$. Show that a is a prime element of R if and only if $(a)_R$ is a prime ideal of R.

Proposition 8.1. *Let R be a commutative ring with identity $1 \neq 0$, and let I be an ideal of R. Then I is prime if and only if R/I is an integral domain.*

Proof. Assume that I is a prime ideal. Then R/I is commutative, because R is, and R/I has identity $1 + I$, which is non-zero because $1 \notin I$. Now suppose that $a, b \in R$ satisfy $(a + I)(b + I) = I$. Then $ab + I = I$, so $ab \in I$, and so either $a \in I$ or $b \in I$ because of the assumption that I is prime. But then either $a + I = I$ or $b + I = I$. Thus, R/I is an integral domain.

Conversely, assume that R/I is an integral domain. In particular, the identity element $1 + I$ is non-zero, so $1 \notin I$, and so $I \neq R$. Now suppose that $a, b \in R$ satisfy $ab \in I$. Then in R/I, $I = ab + I = (a + I)(b + I)$, so because R/I is an integral domain by assumption, it follows that $a + I = I$ or $b + I = I$, which is to say that $a \in I$ or $b \in I$. Thus, I is prime. □

Example. Let us use Proposition 8.1 to show that $x\mathbb{Z}[x]$ is a prime ideal. (There are other ways to show this. See, for example, the first example in Section 6, where we showed that x is a prime element of $\mathbb{Z}[x]$.) Define

$$\begin{aligned} \varphi : \mathbb{Z}[x] &\to \mathbb{Z} \\ f &\mapsto f(0), \end{aligned}$$

a ring homomorphism. Its kernel is

$$\{f \in \mathbb{Z}[x] \mid f(0) = 0\} = \{f \in \mathbb{Z}[x] \mid f \text{ has zero constant term}\} = x\mathbb{Z}[x].$$

Hence, by the First Isomorphism Theorem, $\mathbb{Z}[x]/x\mathbb{Z}[x] \cong \text{Image}(\varphi) = \mathbb{Z}$, an integral domain, so $x\mathbb{Z}[x]$ is a prime ideal in $\mathbb{Z}[x]$.

VI – 9 Maximal ideals

Let R be a commutative ring with identity $1 \neq 0$. An ideal M of R is called *maximal* if $M \neq R$ and the only ideals of R containing M are M and R. (Thus, it is impossible to have an inclusion of the form $M \subsetneqq I \subsetneqq R$ when I is an ideal and M is maximal.)

Proposition 9.1. *Let R be a commutative ring with identity $1 \neq 0$, and let I be an ideal of R. Then I is maximal if and only if R/I is a field.*

Proof. Assume that I is maximal. For the same reasons at the beginning of the proof of Proposition 8.1, R/I is commutative with non-zero identity $1 + I$. Now let $a \in R \smallsetminus I$. We show that $a + I$ has a multiplicative inverse in R/I. Let $J = (\{a\} \cup I)_R$, and note that $J = \{ra + i \mid r \in R \text{ and } i \in I\}$. Because $I \subseteq J$ and I is maximal, either $J = I$ or $J = R$. We cannot have $J = I$, because $a \in J$ but $a \notin I$, so $J = R$. In particular, $1 \in J$, so $1 = ra + i$ for some $r \in R$ and some $i \in I$. Hence, in R/I,

$$1 + I = ra + I = (r + I)(a + I).$$

Thus, R/I is a field.

Conversely, assume that R/I is a field. Then $1 + I \neq I$, so $1 \notin I$, and so $I \neq R$. Now suppose that J is an ideal of R strictly containing I, and choose $b \in J \smallsetminus I$. Then $b + I$ has a multiplicative inverse in the field R/I, i.e., there is $c \in R$ such that

$$1 + I = (c + I)(b + I) = cb + I.$$

Hence, $1 = cb + i$ for some $i \in I$. But $b \in J$ by choice, and $i \in J$ because $I \subseteq J$, so $1 \in J$, and so $J = R$. Thus, I is maximal. □

Corollary 9.2. *If R is a commutative ring with non-zero identity, then every maximal ideal of R is prime.*

Proof. If M is maximal, then R/M is a field and therefore an integral domain, so M is prime by Proposition 8.1. □

Example. We saw in Section 8 that $\mathbb{Z}[x]/x\mathbb{Z}[x] \cong \mathbb{Z}$. Since $\mathbb{Z}$ is an integral domain but not a field, it follows that $x\mathbb{Z}[x]$ is prime but not maximal. Another way to see that $x\mathbb{Z}[x] = (x)_{\mathbb{Z}[x]}$ is not maximal is to observe that $(x)_{\mathbb{Z}[x]} \subsetneqq (2, x)_{\mathbb{Z}[x]} \subsetneqq \mathbb{Z}[x]$.

Example. If p is a prime number, then $\mathbb{Z}/p\mathbb{Z}$ is a field by Proposition 2.2, so $p\mathbb{Z}$ is a maximal ideal of $\mathbb{Z}$.

Example. Let $I = \{f \in \mathbb{Z}[x] \mid f(0) \text{ is even}\}$. Show that I is a maximal ideal of $\mathbb{Z}[x]$ by showing that $\mathbb{Z}[x]/I$ is a field.

Solution: Define

$$\begin{aligned} \varphi : \mathbb{Z}[x] &\to \mathbb{Z}/2\mathbb{Z} \\ f &\mapsto [f(0)]. \end{aligned}$$

We leave it as an exercise to show that φ is a ring homomorphism. Now, φ is surjective, because if $a \in \mathbb{Z}$, then $\varphi(a) = \overline{a}$. Further,

$$\operatorname{Ker}(\varphi) = \{f \in \mathbb{Z}[x] \mid [f(0)] = [0]\} = \{f \in \mathbb{Z}[x] \mid f(0) \text{ is even}\} = I.$$

Therefore, by the First Isomorphism Theorem, $\mathbb{Z}[x]/I \cong \mathbb{Z}/2\mathbb{Z}$, a field. Thus, I is maximal.

Maximal ideals in principal ideal domains

Proposition 9.3. *Let R be a principal ideal domain, and let $a \in R \smallsetminus \{0\}$. Then a is irreducible if and only if $(a)_R$ is maximal.*

Proof. Assume that a is irreducible, and let J be an ideal of R strictly containing $(a)_R$. Because R is a principal ideal domain, $J = (b)_R$ for some $b \in R$. Therefore, there is $c \in R$ such that $a = cb$, and because a is irreducible, either b or c is a unit. If c were a unit, then we would have $(cb)_R = (b)_R$, i.e., $(a)_R = J$, which is not the case. Consequently, b is a unit, so $J = R$. Thus, $(a)_R$ is maximal.

Conversely, if $(a)_R$ is maximal, then it is prime by Corollary 9.2, so a is a prime element of R, and so a is irreducible by Proposition 6.1. □

Example. In $\mathbb{Q}[x]$, the ideal $(x^2-x-5)_{\mathbb{Q}[x]}$ is maximal, because x^2-x-5 is irreducible in $\mathbb{Q}[x]$ (exercise). By contrast, $(x^2 - x - 6)_{\mathbb{Q}[x]}$ is not maximal, because $x^2 - x - 6 = (x + 2)(x - 3)$.

Example. Show that the ring $\mathbb{Z}[i]/3\mathbb{Z}[i]$ is a field, but that $\mathbb{Z}[i]/5\mathbb{Z}[i]$ is not (and is not even an integral domain, in fact).

Solution: We saw in Section 6 that 3 is irreducible in $\mathbb{Z}[i]$, and we pointed out at the end of the same section that $\mathbb{Z}[i]$ is a principal ideal domain. Therefore, by Proposition 9.3, $3\mathbb{Z}[i]$ is a maximal ideal of $\mathbb{Z}[i]$, so $\mathbb{Z}[i]/3\mathbb{Z}[i]$ is a field by Proposition 9.1.

As for 5, observe that $5 = (1 + 2i)(1 - 2i)$. Neither of the factors $1 + 2i$ or $1 - 2i$ is a unit, because each has norm 5, so 5 is not irreducible. Alternatively, use the fact that the only units in $\mathbb{Z}[i]$ are $1, -1, i, -1$ (exercise). Therefore, 5 is not irreducible in $\mathbb{Z}[i]$, so Proposition 6.1 shows that 5 is not a prime element of $\mathbb{Z}[i]$, so $5\mathbb{Z}[i]$ is not a prime ideal, and so $\mathbb{Z}[i]/5\mathbb{Z}[i]$ is not an integral domain.

Another justification that $\mathbb{Z}[i]/5\mathbb{Z}[i]$ is not an integral domain goes as follows. For brevity, if $\alpha \in \mathbb{Z}[i]$, let $[\alpha] = \alpha + 5\mathbb{Z}[i]$. We know from the equation $(1 + 2i)(1 - 2i) = 5$ that $[1+2i][1-2i] = [0]$. We claim that $[1+2i]$ and $[1-2i]$ are non-zero. If $[1+2i] = [0]$, then $1 + 2i = 5\alpha$ for some $\alpha \in \mathbb{Z}[i]$, so taking norms, we have

$$5 = N(1 + 2i) = N(5\alpha) = N(5)N(\alpha) = 25N(\alpha),$$

which is impossible because $25 \nmid 5$. Thus, $[1 + 2i] \neq [0]$. The same argument works for $[1 - 2i]$.

VI – 10 Unique factorization domains

Let R be a commutative ring with identity $1 \neq 0$, and let $a, b \in R$. We will say that a is *associate* to b if there is $u \in R^\times$ such that $a = ub$. Note that, in that case, $b = u^{-1}a$, so b is associate to a as well.

A *unique factorization domain* is an integral domain R such that, for every non-zero $a \in R$ that is not a unit, the following both hold:

(i) a is a product of irreducible elements. (This includes the possibility that a is itself irreducible.)

(ii) The factorization of a into irreducibles is essentially unique, in the sense that if $\pi_1 \cdots \pi_m$ and $\pi'_1 \cdots \pi'_n$ are two such factorizations, then $m = n$ and, after a reordering of the factors if necessary, π_i is associate to π'_i for all $i \in \{1, \ldots, m\}$.

Example. The Fundamental Theorem of Arithmetic, Theorem 3.2 in Section II, shows that $\mathbb{Z}$ is a unique factorization domain, the irreducible elements being $\pm p$ where p is a prime number. For example, while

$$2 \cdot 3 \cdot (-5) \cdot (-3) = p_1 p_2 p_3 p_4$$
$$5 \cdot (-2) \cdot (-3) \cdot 3 = q_1 q_2 q_3 q_4$$

are two factorizations of 90 into irreducibles in $\mathbb{Z}$, they are essentially the same factorization. If $\sim$ denotes the relation "is associate to", then

$$p_1 \sim q_2, \quad p_2 \sim q_4, \quad p_3 \sim q_1, \quad p_4 \sim q_3.$$

Other unique factorization domains besides $\mathbb{Z}$ exist. To see this, we first give a lemma.

Lemma 10.1. *Let R be a principal ideal domain and $I_0 \subseteq I_1 \subseteq I_2 \subseteq \cdots$ a chain of ideals in R. Then there is $n \geq 0$ such that $I_m = I_n$ for all $m \geq n$.*

Proof. Let $I = \bigcup_{m=0}^{\infty} I_m$, an ideal of R (exercise). Because R is a principal ideal domain, there is $a \in R$ such that $I = (a)_R$. Hence, $a \in I_n$ for some $n \geq 0$. Therefore, if $m \geq n$, then $I_n \subseteq I_m \subseteq I = (a)_R \subseteq I_n$, so $I_m = I_n$. □

Proposition 10.2. *Every principal ideal domain is a unique factorization domain.*

Proof. Let R be a principal ideal domain, and let $a \in R$ be non-zero and not a unit. We first show the existence of a factorization of a into irreducible elements. Assume, for a contradiction, that a is not expressible as a product of irreducible elements. In particular, a cannot be irreducible itself, so $a = a_1 a'_1$ where a_1, a'_1 are non-units. If each of a_1 and a'_1 were a product of irreducible elements, then a would be as well. Therefore, one of these, a_1 say, is not a product of irreducible elements. Then a_1 is not irreducible, so $a_1 = a_2 a'_2$ where a_2 and a'_2 are non-units. Repeating this argument, we obtain a sequence $a_0, a_1, a_2, \ldots$, where $a_0 = a$, such that for all $n \geq 0$, a_{n+1} divides a_n but is

not associate to it. Therefore, $(a_0)_R \subsetneqq (a_1)_R \subsetneqq (a_2)_R \subsetneqq \cdots$, contradicting Lemma 10.1. Thus, a is expressible as a product of irreducible elements.

Now we turn to uniqueness. Suppose that

$$\pi_1 \cdots \pi_m = \pi'_1 \cdots \pi'_n, \tag{10.1}$$

where the π_i and π'_j are irreducible. There is no harm in assuming that $m \leq n$. In a principal ideal domain, every irreducible element is prime (Proposition 6.3), so because $\pi_1 \mid \pi'_1 \cdots \pi'_n$, $\pi_1 \mid \pi'_j$ for some j. Without loss of generality, we may assume that $\pi_1 \mid \pi'_1$. But π'_1 is irreducible, so $\pi'_1 = u_1\pi_1$ for some unit u, so π_1 and π'_1 are associate to each other. Cancelling off π_1 from both sides of (10.1), we obtain $\pi_2 \cdots \pi_m = u_1\pi'_2 \cdots \pi'_n$. Continuing in this way (we can make this formal with an induction argument), we eventually arrive at $1 = u_1 \ldots u_m \pi'_{m+1} \cdots \pi'_n$, where $u_1, \ldots, u_m \in R^\times$, and $\pi_i = \pi'_i$ for all $i \in 1, \ldots, m$. If n were greater than m, then $\pi'_{m+1} \cdots \pi'_n$ would be both a unit and a product of one or more irreducible elements, which is not possible. Therefore, $m = n$. □

Let us summarize what we have proven in our study of Euclidean domains, principal ideal domains, and unique factorization domains:

- every field is a Euclidean domain (for trivial reasons),
- every Euclidean domain is a principal ideal domain,
- every principal ideal domain is a unique factorization domain, and
- every unique factorization domain is an integral domain (simply by definition).

In particular, all of the following are unique factorization domains according to the theory we have developed:

- $\mathbb{Z}$ (Euclidean by Proposition 1.1 in Section II)
- $F[x]$ where F is a field (Euclidean by Section 4)
- $\mathbb{Z}[\sqrt{2}]$ (Euclidean by Section 4)
- $\mathbb{Z}[i]$ (Euclidean by a similar proof to that in the case of $\mathbb{Z}[\sqrt{2}]$)

Note that there exist principal ideal domains that are not Euclidean domains. An example is the ring $\{x + y\alpha \in \mathbb{C} \mid x, y \in \mathbb{Z}\}$, where $\alpha = \frac{1}{2}(1 + \sqrt{-19})$.

There are also unique factorization domains that are not principal ideal domains. For example, it is a fact that if R is a unique factorization domain, then so is the polynomial ring $R[x]$. Therefore, $\mathbb{Z}[x]$ is a unique factorization domain that is not a principal ideal domain.

Appendix

Appendix: 1 Proof of Proposition 2.1 in Section I

We recall the statement to be proven:

Let X and Y be non-empty sets and $f : X \to Y$ a function.

(i) f is injective if and only if there is $g : Y \to X$ such that $g \circ f = \mathbf{1}_X$.

(ii) f is surjective if and only if there is $h : Y \to X$ such that $f \circ h = \mathbf{1}_Y$.

(iii) f is bijective if and only if there is $g : Y \to X$ such that $g \circ f = \mathbf{1}_X$ and $f \circ g = \mathbf{1}_Y$.

Proof. (i) If $g \circ f = \mathbf{1}_X$ and $f(x_1) = f(x_2)$, then

$$x_1 = \mathbf{1}_X(x_1) = g \circ f(x_1) = g(f(x_1)) = g(f(x_2)) = g \circ f(x_2) = \mathbf{1}_X(x_2) = x_2,$$

so f is injective.

Conversely, suppose that f is injective. Then for each $y \in \text{Image}(f)$, there is a unique $x_y \in X$ such that $f(x_y) = y$. Now fix any $\tilde{x} \in X$, and define

$$\begin{array}{rcl} g : Y & \to & X \\ y & \mapsto & \begin{cases} x_y & \text{if } y \in \text{Image}(f), \\ \tilde{x} & \text{otherwise.} \end{cases} \end{array}$$

Then $g \circ f = \mathbf{1}_X$.

(ii) If $f \circ h = \mathbf{1}_Y$ and $y \in Y$, then

$$y = \mathbf{1}_Y(y) = f \circ h(y) = f(h(y)) \in \text{Image}(f),$$

so f is surjective.

Conversely, suppose that f is surjective. Then for each $y \in Y$, there is $x_y \in X$ such that $f(x_y) = y$. If we define

$$\begin{array}{rcl} h : Y & \to & X \\ y & \mapsto & x_y, \end{array}$$

then $f \circ h = \mathbf{1}_Y$.

(iii) If $g \circ f = \mathbf{1}_X$ and $f \circ g = \mathbf{1}_Y$, then f is both injective and surjective by (i) and (ii), so it is bijective.

Conversely, suppose that f is bijective. Then by (i) and (ii), there are $g, h : Y \to X$ such that $g \circ f = \mathbf{1}_X$ and $f \circ h = \mathbf{1}_Y$. But

$$g = g \circ \mathbf{1}_Y = g \circ (f \circ h) = (g \circ f) \circ h = \mathbf{1}_X \circ h = h.$$

□

Appendix: 2 Equivalence of the two forms of induction

Proposition 2.1. *The two forms of induction (called "first form" and "second form" in Section I–4) are equivalent, i.e., each implies the other.*

Proof. Assume the first form of induction, let $n_0 \in \mathbb{Z}$, and suppose that we have statements $P(n)$ for $n \geq n_0$ such that

(i) $P(n_0)$ is true, and

(ii) for all $n \geq n_0$, if $P(k)$ is true for all $k \in \{n_0, \ldots, n\}$, then $P(n+1)$ is true.

For each $n \geq n_0$, let $Q(n)$ be the statement "$P(k)$ is true for all $k \in \{n_0, \ldots, n\}$". We will show that $Q(n_0)$ is true and that $Q(n)$ implies $Q(n+1)$ for all $n \geq 0$, from which it follows that $Q(n)$ is true for all $n \geq n_0$ by the first form of induction, and so in particular $P(n)$ is true for all $n \geq n_0$.

That $Q(n_0)$ is true follows immediately from (i). Now let $n \geq n_0$ and assume that $Q(n)$ is true. Therefore, $P(k)$ is true for all $k \in \{n_0, \ldots, n\}$. Then $P(n+1)$ is true by (ii), so $P(k)$ is true for all $k \in \{n_0, \ldots, n+1\}$, which is to say that $Q(n+1)$ is true, as desired.

Conversely, assume the second form of induction, let $n_0 \in \mathbb{Z}$, and suppose that we have statements $P(n)$ for $n \geq n_0$ such that

(i)' $P(n_0)$ is true, and

(ii)' for all $n \geq n_0$, if $P(n)$ is true, then $P(n+1)$ is true.

If $n \geq n_0$ and $P(k)$ is true for all $k \in \{n_0, \ldots, n\}$, then in particular $P(n)$ is true, so $P(n+1)$ is true by (ii)'. Therefore, by the second form of induction, $P(n)$ is true for all $n \geq n_0$, and we are done. □

Appendix: 3 Proof of Proposition 3.1 in Section V

We recall the statement to be proven:

Let $\varphi : R \to S$ be a ring homomorphism. Then the following are equivalent:

(i) φ is an isomorphism.

(ii) φ is bijective (i.e., injective and surjective).

(iii) $\mathrm{Ker}(\varphi) = \{0_R\}$ and $\mathrm{Image}(\varphi) = S$.

Proof. The equivalence of injectivity and the vanishing of the kernel is Proposition 2.4 in Section V, and it is immediate that φ is surjective if and only if $\text{Image}(\varphi) = S$, so (ii) and (iii) are equivalent.

That (i) implies (ii) follows from Proposition 2.1 in Section I. We may complete our proof by showing that (ii) in turn implies (i). Assume that φ is both injective and surjective. By Proposition 2.1 in Section I again, there is a map $\psi : S \to R$ such that $\psi \circ \varphi = \mathbf{1}_R$ and $\varphi \circ \psi = \mathbf{1}_S$. All that remains is to show that ψ is a ring homomorphism. To that end, let $s_1, s_2 \in S$. Then

$$\begin{aligned}
\psi(s_1 + s_2) &= \psi\Big(\varphi \circ \psi(s_1) + \varphi \circ \psi(s_2)\Big) \quad \text{because } \varphi \circ \psi = \mathbf{1}_S \\
&= \psi\Big(\varphi(\psi(s_1)) + \varphi(\psi(s_2))\Big) \\
&= \psi\Big(\varphi(\psi(s_1) + \psi(s_2))\Big) \quad \text{because } \varphi \text{ respects addition} \\
&= \psi \circ \varphi(\psi(s_1) + \psi(s_2)) \\
&= \psi(s_1) + \psi(s_2) \quad \text{because } \psi \circ \varphi = \mathbf{1}_R,
\end{aligned}$$

$$\begin{aligned}
\text{and} \quad \psi(s_1 \cdot s_2) &= \psi\Big(\varphi \circ \psi(s_1) \cdot \varphi \circ \psi(s_2)\Big) \quad \text{because } \varphi \circ \psi = \mathbf{1}_S \\
&= \psi\Big(\varphi(\psi(s_1)) \cdot \varphi(\psi(s_2))\Big) \\
&= \psi\Big(\varphi(\psi(s_1) \cdot \psi(s_2))\Big) \quad \text{because } \varphi \text{ respects multiplication} \\
&= \psi \circ \varphi(\psi(s_1) \cdot \psi(s_2)) \\
&= \psi(s_1) \cdot \psi(s_2) \quad \text{because } \psi \circ \varphi = \mathbf{1}_R.
\end{aligned}$$

□

Appendix: 4 Proof of Eisenstein's Criterion

Before proving Eisenstein's Criterion, Theorem 7.4 in Section VI, we give some useful lemmas.

Lemma 4.1 (Gauss's Lemma)**.** *Let $f \in \mathbb{Z}[x]$. If $f = gh$ where $g, h \in \mathbb{Q}[x]$, then there are $a, b \in \mathbb{Q}$ such that $ab = 1$ and $ag, bh \in \mathbb{Z}[x]$. Therefore, $f = GH$ where $G, H \in \mathbb{Z}[x]$, G is a non-zero rational times g, and H is a non-zero rational times h.*

Proof. We first prove by induction on $n \geq 1$ the following statement: If $f, g, h \in \mathbb{Z}[x]$ satisfy $nf = gh$, then there are $k, l \in \mathbb{Z}_{\geq 1}$ and $\tilde{g}, \tilde{h} \in \mathbb{Z}[x]$ such that $kl = n$, $g = k\tilde{g}$, and $h = l\tilde{h}$. The case $n = 1$ is immediate. Now let $n > 1$, and assume the statement for all smaller values. Suppose that $f, g, h \in \mathbb{Z}[x]$ satisfy $nf = gh$. Because $n > 1$, it has a prime divisor, p say. Reducing both sides of the equation $nf = gh \bmod p$, we obtain $\overline{0} = \overline{g}\,\overline{h}$ in $(\mathbb{Z}/p\mathbb{Z})[x]$, where a bar denotes the image of the given polynomial under the map $\mathbb{Z}[x] \to (\mathbb{Z}/p\mathbb{Z})[x]$. But $(\mathbb{Z}/p\mathbb{Z})[x]$ is an integral domain, as we saw in

Section VI–2, so either $\overline{g} = \overline{0}$ or $\overline{h} = \overline{0}$. Without loss of generality, we may assume that $\overline{g} = \overline{0}$. Then every coefficient of g is divisible by p, so $g = pg_1$ for some $g_1 \in \mathbb{Z}[x]$. Hence, letting $n' = n/p \in \{1, \ldots, n-1\}$, we have $n'f = g_1 h$. We may now apply the inductive hypothesis to n' to deduce the existence of $k, l \in \mathbb{Z}_{\geq 1}$ and $\tilde{g}, \tilde{h} \in \mathbb{Z}[x]$ such that $kl = n'$, $g_1 = k\tilde{g}$, and $h = l\tilde{h}$. But then

$$\begin{aligned} (pk)l &= pn' = n, \\ (pk)\tilde{g} &= pg_1 = g, \\ l\tilde{h} &= h, \end{aligned}$$

and the induction is complete.

To finish the proof of the lemma, we just observe that if $f = gh$ where $g, h \in \mathbb{Q}[x]$, then there are $c, d \in \mathbb{Z}_{\geq 1}$ such that $cg, dh \in \mathbb{Z}[x]$, and then $nf = (cg)(dh)$ where $n = cd$. Therefore, by the above, there are $k, l \in \mathbb{Z}_{\geq 1}$ such that $kl = cd$ and $\frac{c}{k}g, \frac{d}{l}h \in \mathbb{Z}[x]$. □

Lemma 4.2. *If $f \in \mathbb{Z}[x]$ is a monic polynomial, then it is irreducible in $\mathbb{Q}[x]$ if and only if it is irreducible in $\mathbb{Z}[x]$.*

Proof. Assume first that f is irreducible in $\mathbb{Q}[x]$. In particular, $\deg(f) \geq 1$, so f is not a unit in $\mathbb{Z}[x]$. Now suppose that $f = gh$ where $g, h \in \mathbb{Z}[x]$ and $g \notin \mathbb{Z}[x]^\times$. We show that $h \in \mathbb{Z}[x]^\times$. Let a be the leading coefficient of g. The factorization $f = gh$ shows that $a \in \{1, -1\}$, so if g had degree 0, it would be a unit, contradicting the assumption on g. Therefore, $\deg(g) \geq 1$. But then $g \notin \mathbb{Q}[x]^\times$, so the fact that f is irreducible in $\mathbb{Q}[x]$ implies that $h \in \mathbb{Q}[x]^\times$, i.e., $h = b$ for some $b \in \mathbb{Q}^\times$. But $h \in \mathbb{Z}[x]$, so $b \in \mathbb{Z}$, and then the fact that $ab = 1$ implies that $b \in \{1, -1\} = \mathbb{Z}[x]^\times$, as desired.

Conversely, assume that f is irreducible in $\mathbb{Z}[x]$. Note that if f were a unit in $\mathbb{Q}[x]$, it would have degree 0, and therefore, being monic, would have to be equal to 1, contradicting the assumption that f is irreducible in $\mathbb{Z}[x]$. Thus, f is not a unit in $\mathbb{Q}[x]$. Now suppose that $f = gh$ where $g, h \in \mathbb{Q}[x]$ and $g \notin \mathbb{Q}[x]^\times$. We show that $h \in \mathbb{Q}[x]^\times$. If it were not, then it would have degree at least 1, and then Lemma 4.1 would imply that $f = GH$ for some $G, H \in \mathbb{Z}[x]$ both of degree at least one, contradicting the assumption that f is irreducible in $\mathbb{Z}[x]$. Therefore, $h \in \mathbb{Q}[x]^\times$, as claimed. □

Lemma 4.3. *Let R be an integral domain, and suppose that $g, h \in R[x]$ both have degree at least 1 and satisfy $gh = x^t$ for some $t \geq 2$. Then the constant terms of g and h are both zero.*

Proof. Write

$$g = \sum_{i=0}^{m} a_i x^i \quad \text{and} \quad h = \sum_{j=0}^{n} b_j x^j,$$

where $a_m, b_n \neq 0$. By assumption, $\sum_{i=0}^{k} a_i b_{k-i} = 0$ for all $k \in \{0, \ldots, m+n-1\}$. The case $k = 0$ says that $a_0 b_0 = 0$, so $a_0 = 0$ or $b_0 = 0$. Without loss of generality, we may assume that $a_0 = 0$. We show that, in fact, $b_0 = 0$ as well. Suppose not. Then because $m + n - 1 \geq 1 + 1 - 1 = 1$, we may turn to the case $k = 1$ to see that $a_0 b_1 + a_1 b_0 = 0$,

which implies, along with the fact that $a_0 = 0$ and the assumption that $b_0 \neq 0$, that $a_1 = 0$. But then the case $k = 2$ gives $a_2 = 0$ for similar reasons, and so on. But this will eventually contradict the fact that $a_m \neq 0$. Therefore, $b_0 = 0$ as well. □

We may now prove Eisenstein's Criterion. We recall the statement to be proven:

> Let $f = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in \mathbb{Z}[x]$ be a monic polynomial of degree at least 1, and suppose that there is a prime number p such that $p \mid a_i$ for all $i \in \{0, \ldots, n-1\}$ but $p^2 \nmid a_0$. Then f is irreducible in both $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$.

Proof. By Lemma 4.2, it is enough to show that f is irreducible in $\mathbb{Z}[x]$.

Note that, by assumption, f has positive degree and therefore is not zero or a unit in $\mathbb{Z}[x]$. Suppose now that $f = gh$ where $g, h \in \mathbb{Z}[x]$ and $g \notin \mathbb{Z}[x]^\times$. We show that $h \in \mathbb{Z}[x]^\times$. If $\overline{g}$ and $\overline{h}$ are the reductions of g and h mod p, then the assumption on the coefficients of f implies that

$$x^n = \overline{g}\,\overline{h}$$

in $(\mathbb{Z}/p\mathbb{Z})[x]$. Because $g \notin \mathbb{Z}[x]^\times$ and its leading coefficient is ± 1, its degree is at least 1 and so is the degree of $\overline{g}$. Now, if $\deg(h) \geq 1$, then $\deg(\overline{h}) \geq 1$ as well by the foregoing reasoning, but then Lemma 4.3 would imply that the constant terms of $\overline{g}$ and $\overline{h}$ were both zero in $\mathbb{Z}/p\mathbb{Z}$, which would say that the constant terms of g and h were both divisible by p, which would in turn imply that $p^2 \mid a_0$, contradicting our assumption on a_0. Thus, $\deg(h) = 0$, so $h = b$ for some $b \in \mathbb{Z}$, so because $f = gh$ and f is monic, we must have $h \in \{1, -1\} = \mathbb{Z}[x]^\times$, as desired. □