Algebra I
Section 2: The System of Integers

2.1 Axiomatic definition of Integers.

The first algebraic system we encounter is the **integers.** In this note we list the axioms that determine the system of integers, along with many simple consequences of those axioms. Most of those consequences will be stated without proof, or left as exercises; our main purpose in this section is to survey the facts about the integers you can safely assume in later discussions. Besides, the missing proofs will be handled later on in a more general context (*the theory of rings*).

The integers are a system $(\mathbb{Z}, +, \cdot\,)$ which consists of a set $\mathbb{Z}$ equipped with two operations $(+)$ and $(\cdot)$ that map $\mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{Z}$. We now identify the properties these operations must possess to become the familiar system of integers. At first we shall consider a more general set R equipped with two operations $(+)$ and $(\cdot)$ from $R \times R \rightarrow R$. We will impose axiomatic conditions in stages, finally arriving at the axioms characteristic of the system of integers.

2.1.1 Axioms I: Commutative Ring. The system $(R, +, \cdot\,)$ is a **commutative ring with identity** if the operations have the following properties.

A.1. $(x + y) + z = x + (y + z)$ (associativity of addition)

A.2. $x + y = y + x$ (commutativity of addition)

A.3. There exists an element $0 \in R$ such that $0 + x = x = x + 0$ for all $x \in R$. This is the "zero element" of the system.

A.4. Every element $x \in R$ has an "additive inverse," denoted by $-x$, which has the property $x + (-x) = 0 = (-x) + x$.

Later on we will see that this set of axioms, which govern the $(+)$ operation only, makes $(R, +)$ into what algebraists would call a *commutative group.* We now add some axioms concerning multiplication and its interaction with addition.

M.1. $(x \cdot y) \cdot z = x \cdot (y \cdot z)$ (associativity of multiplication)

M.2. $x \cdot y = y \cdot x$ (commutativity of multiplication)

M.3. There exists in R a "multiplicative identity element," denoted by 1 (or sometimes 1_R), which has the characteristic property that $1 \cdot x = x = x \cdot 1$ for all $x \in R$.

M.4. $x \cdot (y + z) = x \cdot y + x \cdot z$ (distributive law)

M.4' $(x + y) \cdot z = x \cdot z + y \cdot z$ (distributive law)

M.5. $1 \neq 0$ (we exclude the "trivial ring," which has 0 as its only element) □

Below we list important consequences of this set of axioms. You will recognize many familiar attributes of the integers, but be aware that every one of these ancillary properties *must be proved from the fundamental properties listed in Axioms I.* This is not always a simple matter. It is suggested that you try making your own proofs for the items marked with $(*)$ in the following list.

2.1.2 Consequences of Axioms I. For a commutative ring with identity $(R, +, \cdot\,)$ we have

1. The xero element 0 is unique; if $0'$ is any other element in R such that $x + 0 = x = 0 + x$ for all $x \in R$, then $0' = 0$.

   PROOF: Look at $0 + 0'$. By A.3 we get $0' = 0 + 0' = 0$. □

2. Likewise, the additive inverse $-x$ is also unique: if x is fixed and u is any other element in R such that $u + x = 0 = x + u$, then $u = -x$.

   PROOF: Here "$-x$" is any element u such that $u + x = 0 = x + u$. If u, u' are two such elements, look at the combination $(u' + x) + u = u' + (x + u)$. We get

   $$u = 0 + u = (u' + x) + u = u' + (x + u) = u' + 0 = u'$$

   as required. □

3.* The multiplicative identity element 1 is unique: if $1'$ is any other element in R such that $1' \cdot x = x = x \cdot 1'$ for all x, then $1' = 1$.

4.* Consider the "additive inverse map" given by $J(x) = -x$. If you apply this map twice you get back where you started: $-(-x) = x$ for all $x \in R$.
   Hint: Think of "$-x$" as the unique element that makes $\square + x = 0 = x + \square$ when placed in the empty box $\square$. What can you put in the box to make $\square + (-1) = 0$?

5.* $0 \cdot x = 0$ for all $x \in R$.

PROOF: This last is *not* an axiom! Its proof from the axioms is also tricky. Start with the identity $0 + 0 = 0$ (by A.3) and observe that

$$0 \cdot x \;=\; (0 + 0) \cdot x \;=\; 0 \cdot x + 0 \cdot x \quad \text{(by M.4)}$$

which implies

$$\begin{aligned}
0 &= 0 \cdot x + (-(0 \cdot x)) \\
&= \big[0 \cdot x + (0 \cdot x)\big] + (-(0 \cdot x)) \quad \text{(apply previous identity)} \\
&= 0 \cdot x + \big[(0 \cdot x) + (-(0 \cdot x))\big] \quad \text{(by A.1)} \\
&= 0 \cdot x + 0 \quad \text{(by A.3)} \\
&= 0 \cdot x \quad \square
\end{aligned}$$

6.* If -1 is the additive inverse of 1, then $-x = (-1) \cdot x$ for any $x \in R$.
Hint: See Hint to #4.

Moral: Once you have identified the additive inverse -1 of the identity element you can find $-x$ for any other element; just multiply x by (-1).

2.1.3 Exercise. Assuming the properties (1.) – (6.) have already been established, give a proof of the following crucial fact: In any commutative ring,

$$(1) \qquad (-1)^2 = 1$$

where -1 is the additive inverse of the identity element 1.
Hint: The results posted in #4 and #6 might be useful here. □

The following properties follow immediately from (1).

7.* $(-x) \cdot y = x \cdot (-y) = -(x \cdot y)$ for all $x, y \in R$

8.* $(-x) \cdot (-y) = x \cdot y$ for all $x, y \in R$. □

For simplicity, we write $x + (-y) = x - y$ and $x \cdot y = xy$ hereafter. This should cause no confusion.

The rules in Axioms I fall far short of characterizing the integers. The following algebraic system satisfies all these conditions but could hardly be confused with the integers.

2.1.4 Exercise. Let $R = \{0, 1\}$ be a two-element set with two operation $(+)$ and $(\cdot)$ defined by setting

$$1+1=0 \qquad 1+0=0+1=1 \qquad 0+0=0$$

$$1\cdot 1=1 \qquad 1\cdot 0=0\cdot 1=0 \qquad 0\cdot 0=0$$

Verify that $(R,+,\cdot\,)$ satisfies Axioms I. □

One thing missing from Axioms I is any mention of "order properties" such as $x < y$. Since order is an important attribute of the system of integers it's not surprising that Axioms I are not enough. What is surprising is that the order properties of the integers, which most people think of as geometric properties, can be described entirely in algebraic terms.

2.1.5 Axioms II: Order Axioms. Let $(R,+,\cdot\,)$ be a commutative ring with identity $1 = 1_{R}$. This system is said to be an **ordered ring** if there exists a subset $P = P_{R} \subseteq R$ with the following properties.

O.1. R is a *disjoint* union $R = -P \cup \{0\} \cup P$ of the sets P, $-P = \{-x : x \in P\}$, and the zero element.

Notice that $x \in P \Rightarrow -x \in -P$ by definition of $-P$. But we also have $x \in -P \Rightarrow -x \in P$, essentially because $-x = (-1)\cdot x$ and $(-1)^2 = 1$. In fact, $x \in -P \Rightarrow \exists u \in P$ such that $x = -u = (-1)\cdot u$, and hence $-x = (-1)\cdot x = (-1)^2 \cdot u = u \in P$ as claimed. Thus $x \in P \Leftrightarrow -x \in P$.

O.2. $P + P = \{x+y : x, y \in P\} \subseteq P$ ("positive + positive = positive")

O.3. $P \cdot P = \{x\cdot y : x, y \in P\} \subseteq P$ ("positive $\cdot$ positive = positive")

We write $x > 0$ if $x \in P$, $x < 0$ if $x \in -P$, and we define the symbol $x > y$ to mean that $x - y > 0$. Points in the set P are referred to as *positive elements* in R; points in $-P$ are the *negative elements.* □

Many important properties must now be derived. We omit most discussion, but you might try proving the items marked by $*$ on your own. (If you do, assume all previous statements have been proved, so you can use them in dealing with the problem at hand.)

2.1.6 Consequences of the Order Axioms. Any system $(R,+,\cdot\,,>)$ satisfying Axioms I-II has the following properties.

1. Given an $x \in R$ exactly one of the following possibilities is true: (i) $x < 0$, (ii) $x = 0$, (iii) $x > 0$.
2. $1 > 0$

Statement (1.) recapitulates Axiom O.1. Statement (2.) is so important, and its proof is such a good illustration of the interplay between algebraic and order axioms, that we give the details.

> PROOF OF (2): By (M.5) we know that $1 \neq 0$. In view of the "trichotomy" property (1.) we must have $1 < 0$ if we don't have the desired property $1 > 0$. As noted in 2.1.5, we would then have $-1 > 0$ (by our comments on O.1). However, we have already seen in 2.1.3 that $(-1)^2 = 1$; hence (by O.2) we get $1 > 0$, and cannot have $1 < 0$. Conclusion: the only viable possibility is that $1 > 0$, as claimed. □

Returning to our list of consequences, we have

3. $x > 0$ and $y > 0 \Longrightarrow xy > 0$ and $x + y > 0$ (a recap of O.2,O.3)

4.* $x > y \Longrightarrow x + c > y + c$, for all $c \in R$

This says: an inequality remains valid if you add *any number* to both sides.

5.* $x > y$ and $y > z \Longrightarrow x > z$ (transitivity of the order relation)

6. $x \neq 0 \Longrightarrow x^2 > 0$

7. $c > 0$ and $x > y \Longrightarrow xc > yc$ and $c < 0 \Longrightarrow xc < yc$ (reverses inequality)

This says: an inequality remains valid if you multiply both sides by a *positive* number.

8. The usual "rules of signs" hold:

$$\begin{array}{ll} x > 0 \text{ and } y > 0 \Rightarrow xy > 0 & \text{– i.e. } (+)\cdot(+) = (+) \\ x > 0 \text{ and } y < 0 \Rightarrow xy < 0 & \text{– i.e. } (+)\cdot(-) = (-) \\ x < 0 \text{ and } y < 0 \Rightarrow xy > 0 & \text{– i.e. } (-)\cdot(-) = (+) \end{array}$$

9. $x > y \Longleftrightarrow -x < -y$ □

$*$**2.1.7 Proposition.** *If a system* $(R, +, \cdot, >)$ *satisfies Axioms I-II, prove that*

(a) No Zero Divisors*: If* $xy = 0$*, then either* $x = 0$ *or* $y = 0$ *(or both).*

(b) Cancellation Law*: If* $a \neq 0, ax = ay$*, then* $x = y$*.*

Proof: Since $a \cdot 0 = 0$ we may multiply both sides by (-1) as necessary to make $x, y \geq 0$. Obviously the product is zero if either x or y is zero. Otherwise, we have $x, y > 0$ and then $xy > 0$ by Axiom O.3.

For (b), we have $ax = ay \Leftrightarrow 0 = ax - ay = a(x - y)$ and since $a \neq 0$ we must have $x - y = 0$ or $x = y$, as required. □ $(x > 0, x = 0, x < 0)$ and the sign of y. □

2.1.8 Exercise. In an ordered ring $(R, +, \cdot, >)$ suppose we have $a > b$ and $c > d$.

(a) Do we have $ac > bd$? Prove or give a counterexample.

(b) Suppose $a > b$ *and* we assume that $a > 0, b > 0$. If we do not impose any conditions on c, d except $c > d$, do we get $ac > bd$? □

2.1.9 Exercise. In an ordered ring, show that

$$\text{If } a > 0 \text{ and } b > 0 \text{ then } a > b \Longleftrightarrow a^2 > b^2 \qquad (2)$$

Hint: $(b^2 - a^2) = (b - a)(b + a)$. Use Rule of Signs. □

There are, still, many algebraic systems satisfying Axioms I-II, for example the rational numbers $\mathbb{Q}$, or the real numbers $\mathbb{R}$. To exclude those, we introduce what is probably the most distinctive (and subtle) property of the integers. This axiom specifies a connection between the set $P_R = \{x \in R : x > 0\}$ of positive elements in an ordered ring $(R, +, \cdot, >)$ and the "counting numbers" in R, which are obtained by starting with the identity element 1 and repeatedly forming "successors"

Start 1
1+1 (which we label "2")
1+1+1 = 2 + 1 (which we label "3")
⋮
etc.

The distinctive feature of the system of integers is that this process yields all the positive elements. The idea is to make this intuitive statement into our third and last axiom.

Unfortunately, the phrase "... etc." used above is not a proper definition of the "counting numbers" because the suggested construction cannot be completed in a finite amount of time. Here's one way to craft a correct definition of the counting numbers in any system that satisfies Axioms I. The definition is "existential" rather than "constructive."

2.1.10 Definition. *A subset* $S \subseteq P_R$ *in the set of positive elements in a commutative ordered ring* R *is an* **inductive set** *if*

(i) *The identity element 1 lies in S:* $1 \in S$

(ii) *The* **successor** $s+1$ *of any element in S is also in S:* $s \in S \implies s+1 \in S$.

Obviously P_R itself is an inductive set, but in some systems there are much smaller inductive sets than P_R – i.e. the element 1 and its successors fail to "generate" the full set of positive elements in R (the system $\mathbb{R}$ is one example). Our final axiom forces the counting numbers to have the properties we intuitively expect of them. This axiom turns out to be the final step in characterizing the system of integers.

2.1.11 Axiom III: The Induction Axiom. *We assume that $(R, +, \cdot, >)$ is an ordered commutative ring (a system satisfying* Axioms I-II*) and then require that the set of positive elements $P_R = \{x \in R : x > 0\}$ has the following property.*

Induction Property: *If S is a subset of P_R has the properties*

(a) $1 \in S$

(b) $s \in S \Longrightarrow s+1 \in S$ (*i.e. "if s lies in S, so does its* successor $s+1$")

then S is equal to the full set P_R of positive elements in R. □

It can be proved that there exists a *unique* algebraic system $(R, +, \cdot, >)$ satisfying Axioms I-III. The proof is complicated, and we will not go into the details here. We call the resulting system the **system of integers**, and hereafter we denote it by the symbol $\mathbb{Z}$ instead of R; we will also use the traditional notation $\mathbb{N}$ for the counting numbers $\{1, 2, \ldots\}$ instead of P_R. Notice that $\mathbb{Z} = -\mathbb{N} \cup \{0\} \cup \mathbb{N}$ (disjoint union), by Axiom II.

Below we list some useful properties of the integers that may seem intuitively obvious, but which must be proved from Axioms I-III and their previously established consequences.

2.1.12 Exercise. Prove that the identity element 1 is the *smallest* number in $\mathbb{N}$. – i.e. that $1 \leq n$ for all $n \in \mathbb{N}$.
Hint: Consider the set $S = \{n \in \mathbb{N} : n \geq 1\}$ and show it has properties (a),(b) in the Induction Axiom. □

2.1.13 Exercise. If $n \in \mathbb{N}$ and $n > 1$ prove that $n - 1 \in \mathbb{N}$.
Hint: In $\mathbb{Z}$ we have $\mathbb{N} = P$. Use 2.1.12. □

2.1.14 Exercise. If $n \in \mathbb{N}$ and $n > 1$, prove that $n \geq 2$ – i.e. there are no natural numbers between 1 and its successor $2 = 1 + 1$. □

2.1.15 Exercise. If $n \in \mathbb{N}$ show that there cannot be an $x \in \mathbb{N}$ such that $n < x < n + 1$. □

2.1.16 Exercise. If $m, n \in \mathbb{N}$ and $m > n$, prove that $m - n \in \mathbb{N}$. □

We list these small results because they have a tendency to come up often. You are urged to work out proofs for Exercises 2.1.12-16; they aren't very hard.

Next we come to a really important fact. We relegate the somewhat intricate proof to Appendix A at the end of this chapter.

2.1.17 Theorem (The Minimum Property). *If $S \subseteq \mathbb{N}$ is non-empty, then there exists a unique minimum element in S – an element $s_0 = \min\{S\}$ in S such that $s_0 \leq s$ for all $s \in S$.*

The minimum element s_0 is obviously unique, once we know it exists.

The Minimum Property is sometimes taken as an alternative Axiom III$'$ in place of Axiom III.

(3) *Equivalence of Induction and Minimum Properties:* If $(R, +, \cdot, >)$ is any system satisfying Axioms I-II, then (Axiom III) holds in this system if and only if (Axiom III$'$) holds.

Theorem 2.1.17 above is just the implication (Axiom III)$\Rightarrow$(Axiom III$'$). For completeness a full proof of this equivalence is given in Appendix A.

We have emphasized the Induction Principle over the Minimum Principle because induction is the foundation of all mathematics as we know it. The following rewording of Axiom III shows how this principle gets used in practice.

2.1.18 The Induction Principle in Practice. Suppose an assertion $P(n)$ has been assigned to each counting number $n \in \mathbb{N}$, and each statement is either true or false. Suppose we can show that

(a) Statement $P(1)$ is true.

(b) *If* statement $P(n)$ is assumed to be true, we can then prove $P(n+1)$ true based on this information. (In symbolic logic shorthand: $P(n) \Rightarrow P(n+1)$ for all $n \in \mathbb{N}$.)

Conclusion: the statement $P(n)$ must be true for *all* $n \in \mathbb{N}$.

Discussion: This follows from Axiom III by looking at the set $S = \{n \in \mathbb{N} : P(n) \text{ is true}\}$. We have $1 \in S$ by (a), and (b) tells us that S is "closed" under formation of successors, so that $s \in S \Rightarrow s+1 \in S$. Thus $S = \mathbb{N}$. Notice that by appealing to the Induction Axiom we are able to establish the truth of *infinitely many different statements* $P(1), P(2), \ldots$ with a finite amount of effort! □

2.1.19 Example. Show that for any $n \in \mathbb{N}$

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2} \qquad \text{for all } n \in \mathbb{N} \tag{4}$$

Here $1 + 2 + \ldots + n$ means "the sum of all integers k such that $1 \leq k \leq n$."

Discussion: Consider the statements $P(n) =$ (the identity (4) is true). Obviously $P(1)$ is true because 1 is the only integer k such that $1 \leq k \leq 1$, and $1 = (2 \cdot 1)/2$. Now suppose we know $P(n)$ is true for some n. Then the next statement $P(n+1)$ can be rewritten

$$\begin{aligned} (1 + 2 + \ldots + n) + (n+1) &= \frac{n(n+1)}{2} + (n+1) \qquad \text{(since } P(n) \text{ is true)} \\ &= \frac{n(n+1) + 2(n+1)}{2} \\ &= \frac{(n+2)(n+1)}{2} = \frac{(n+1)[(n+1)+1]}{2} \quad , \end{aligned}$$

and this is just what we need to validate $P(n+1)$. By induction, $P(n)$ is true for all n. □

2.1.20 Exercise. Prove that

$$1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

is true for all $n \in \mathbb{N}$. □

***2.1.21 Exercise.** Prove that the sum of the first n odd integers is equal to n^2

$$n^2 = \sum_{k=1}^{n} (2k-1) = 1 + 3 + \ \ldots \ + (2n-1)$$

for all $n \in \mathbb{N}$. □

2.1.22 Exercise. Prove that

$$\sum_{k=1}^{n} k^3 = \big(\sum_{k=1}^{n} k\big)^2 = (1+2+\ldots+n)^2 = \frac{n^2(n+1)^2}{4}$$

for all $n \in \mathbb{N}$.
Hint: Use 2.1.19. □

The following example shown how an "argument by induction" can go astray if you are not meticulous in checking that your Induction Hypothesis actually works when you claim it does. The biggest culprit is careless language, as below.

2.1.23 Example (A Paradox?). We claim that in any finite set of billiard balls, all members of the set have the same color.

"Proof": The inductive hypothesis is:

> *Inductive Hypothesis* $P(n)$: For any counting number $n \in \mathbb{N}$ and any set of billiard balls with n members, all the balls have the same color

This is certainly true for $n = 1$: there is just one ball. Now suppose $n > 1$ and we know that $P(n)$ is true. We prove that $P(n+1)$ must be true, and hence that $P(n)$ is true for all counting numbers, as follows.

Consider any set A of $n + 1$ balls, and the subsets

$$B = (\text{first } n \text{ balls}) \qquad C = (\text{last } n \text{ balls})$$

The inductive hypothesis applies to both B and C, so all balls in B have the same color, and likewise for the balls in C. Since the two sets have a ball in common, all the balls in their union $A = B \cup C$ have the same color, proving that $P(n + 1)$ is true. By the Induction Axiom, $P(n)$ is true for all counting numbers n so *all* billiard balls have the same color. □

The conculsion is absurd. Can you spot the error in this "proof"?

2.1.24 Definition (Finite sets). *Two sets A, B are said to have the* **same cardinality***, or size, if there exists a bijection $\psi : A \to B$; this is indicated by writing $A \approx B$. This relationship means that the points in A can be matched exactly with those in B. Of course there may be many ways to achieve the match, so ψ is not unique. It is the existence of some ψ that makes $A \approx B$*

In $\mathbb{N}$ we define an **interval** $[1, n] = \{k \in \mathbb{N} : 1 \leq k \leq n\}$ *for each $n \in \mathbb{N}$, and we say that a set A is* **finite** *if there exists some n such that $A \approx [1, n]$. If A is not finite we say that A is* **infinite***. The two possibilities are indicated by writing $|A| = n$ or $|A| = \infty$ respectively. The empty set is also regarded as a finite set, with cardinality $|\emptyset| = 0$.*

Our definition of $|A| = n$ attempts to capture what we mean when we say that "A is finite and contains exactly n points." However, there is a long road from these basic definitions to proofs that the relations $|A| = n$ and $|A| = \infty$ have the properties we intuitively expect of them. Everything must in the end be proved from the axioms governing the integers. We will not go into those details in these *Notes*, except to say that virtuoso use of the Induction Axiom is needed to verify such "intuitively obvious" properties as:

1. If $|A| = m$ and $|B| = n$ then $A \approx B \Leftrightarrow m = n$. In particular, there cannot be a bijection between $[1, m]$ and $[1, n]$ unless $m = n$.

Thus each finite set is associated with a uniquely defined natural number n, its **cardinality**.

2. If A is finite and $B \subseteq A$ then B is finite and $|B| \leq |A|$, with equality if and only if $A = B$. Subsets of an infinite set can be finite or infinite.
3. *The Arithmetic of Cardinalities.* If A and B are *disjoint* finite sets, then $A \cup B$ is finite and $|A \cup B| = |A| + |B|$. If the sets are not disjoint, $A \cup B$ is still finite but all we can say about its cardinality is that $|A \cup B| \leq |A| + |B|$. More generally, suppose we have a finite collection of sets, say $\{A_k : k \in [1, n]\}$ for some $n \in \mathbb{N}$, such that each A_k is finite. Then their union $\bigcup_{k \in [1,n]} A_k = \bigcup_{k=1}^{n} A_k$ is also finite, and

   $$\Big| \bigcup_{k=1}^{n} A_k \Big| \leq \sum_{k=1}^{n} |A_k| \quad .$$

   Equality holds if the sets are pairwise disjoint in the sense that $A_i \cap A_j = \emptyset$ for all $i \neq j$.
4. If A is finite and B is infinite there cannot be a bijective map between them.
5. If A is an infinite set and B a proper subset, so $B \subseteq A$, $B \neq A$, it is possible that $B \approx A$. Thus an infinite set can be equivalent to a proper subset of itself. Some examples: $\mathbb{Z} \approx \mathbb{N}$, $\mathbb{Z} \approx 2\mathbb{Z} = \{$ all even integers $\} = \{2k : k \in \mathbb{Z}\}$ and $\mathbb{Z} \approx \mathbb{Q} = \{$ all rational numbers $\}$.
6. There are infinite sets A, B that do not have the same size – i.e. A and B are both infinite, but there is no bijection between them.

A famed theorem of Cantor showed that the infinite sets $\mathbb{Z}$ and $\mathbb{R}$ are not of the same cardinality, and hence represent different "orders of infinity." It was also shown that the real line $\mathbb{R}$ *has the same cardinality as any Euclidean space* $\mathbb{R}^n$, and in particular that the line $\mathbb{T}$ and the plane $\mathbb{R}^2 = \mathbb{R} \times \mathbb{T}$ are equivalent, so $\mathbb{R} \approx \mathbb{R}^2$.

7. If A is a countable infinite set and $B \subseteq A$ then B is either finite or countable. There can be no orders of infinity lurking between "finite" and "countable."
8. *Schroder-Bernstein Theorem.* Let A, B be any sets, finite or not, and suppose that

   (i) $A \approx B'$ for some subset $B' \subseteq B$

   (ii) $B \approx A'$ for some subset $A' \subseteq A$

   Then there exists a bijection $A \approx B$.
   Note: In fact ϕ can be constructed explicitly in countably many recursive steps by slicing and dicing the original maps f, g and suitably reassembling the pieces.

The last theorem is nontrivial even when the sets A, B are finite. A brief self-contained proof is given in Appendix B.

The next (and last) basic property of finite sets is ultimately a consequence of the Minimum Property 2.1.17. It tells you how to identify the finite subsets of the integers, and states a property that is often invoked for such sets in place of 2.1.17.

9. A subset $S \subseteq \mathbb{Z}$ is finite if and only if S lies in some interval in the integers $[a, b] = \{x \in \mathbb{Z} : a \leq x \leq b\}$ with $a \leq b$. To put it differently, a finite set is one that is bounded from above and from below. A nonempty *finite* set S in the integers has a largest and a smallest element – i.e. there exist elements $a, b \in S$ such that $a \leq s \leq b$ for all $s \in S$.

As indicated above, we forego the details of proof (which are quite involved) since this is not a course in set theory. However, we will make free use of the properties (1.)-(9.) as needed.

2.1.25 Exercise. Suppose a nonempty set A is equivalent to the interval $[1, m]$ in $\mathbb{N}$ and is also equivalent to $[1, n]$. Prove that $m = n$, so the cardinality $n = |A|$ is well defined.
Hints: This amounts to proving that $[1, m] \approx [1, n] \Rightarrow m = n$. We may assume $m \geq n$; procede by induction on m. You are given a bijection $f : [1, n] \to [1, m]$ which may have to be modified during your discussion.

2.1.26 Exercise. If A is a nonempty finite subset of $\mathbb{Z}$ prove that A is bounded from above and from below – i.e. there exist $a, b \in \mathbb{Z}$ such that $a \leq x \leq b$ for all $x \in A$.
Hint: Since $A \approx [1, n]$ (if it is not empty) we can list its elements as $a_1, \ldots, a_i, \ldots, a_n$. Try some combinations of $|a_1|, \ldots, |a_n|$ (or $-|a_1|, \ldots, -|a_n|$). $\square$

***2.1.27 Exercise.** If A is a nonempty finite subset of $\mathbb{Z}$ prove that A has a largest and a smallest element.
Hints: Start with 2.1.26. If $b \in \mathbb{Z}$ what happens to $\min(A)$ and $\max(A)$ when you replace A by the translated set $b + A = \{b + a : a \in A\}$? What happens when you replace A by $-A = \{-a : a \in A\} = (-1) \cdot A$? You will of course have to invoke the Minimum Principle. $\square$

***2.1.28 Exercise.** Show that the following infinite sets have the same cardinality by finding explicit bijections between them.

(a) $\mathbb{Z}$ and the set of even integers $2\mathbb{Z} = \{2n : n \in \mathbb{Z}\}$.

(b) The sets $2\mathbb{Z}$ and $2\mathbb{Z} + 1$ of even and odd integers.

(c) The sets $\mathbb{N}$ and $A = \{n \in \mathbb{Z} : n \leq 50\}$

(d) The sets $\mathbb{N}$ and $\mathbb{Z}_+ = \{0\} \cup \mathbb{N} = \{n \in \mathbb{Z} : n \geq 0\}$

(e) The real line $\mathbb{R}$ and the unbounded interval $(0, +\infty) = \{x \in \mathbb{R} : x > 0\}$.

(f) The real line $\mathbb{R}$ and the bounded interval $(0, 1) = \{x \in \mathbb{R} : 0 < x < 1\}$. $\square$

***2.1.29 Exercise.** Using the Schroeder-Bernstein Theorem, prove that the following sets must have the same cardinality by producing explicit one-to-one (but not necessarily surjective) maps $f : A \to B$ and $g : B \to A$.

(a) The real line $\mathbb{R}$ and the unbounded interval $[0, +\infty) = \{x \in \mathbb{R} : x \geq 0\}$.

(b) The real line $\mathbb{R}$ and the bounded interval $[-1, 1] = \{x \in \mathbb{R} : -1 \leq x \leq 1\}$.

(c) The intervals $[0, 1], [0, 1)$, and $(0, 1)$ in the real line.

(d) The unit disc in the plane $A = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$ and the unit square $B = \{(x, y) \in \mathbb{R}^2 : -1 \leq x, y \leq 1\}$.

(e) The unit disc in the plane $A = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$ and the entire plane $B = \mathbb{R}^2$. $\square$

2.2 Arithmetic in the Integers $\mathbb{Z}$.

This is still a quick survey, but now we will give proofs more often. Once we hit Section 2.3 *all* results will be proved from the axioms or previously established consequences. We begin with the properties of absolute value $|a|$.

2.2.1 Definition. *The* **absolute value** *of $x \in \mathbb{Z}$ is*

$$|x| = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -x & \text{if } x < 0 \end{cases}$$

The absolute value can be viewed as an operation that maps $\mathbb{Z}$ into $\mathbb{Z}$. Its range is the set $\mathbb{Z}_+ = \{n \in \mathbb{Z} : n \geq 0\} = \mathbb{N} \cup \{0\}$ of nonnegative integers.

The following properties of absolute value are immediate consequences of this definition.

1. $|x| \geq 0$ and $|x| = 0 \Leftrightarrow x = 0$.
2. $|x| = +x$ or $-x$ for any $x \in R$.
3. $|-x| = |x|$ for all $x \in R$
4. $|x| \geq x$ and $|x| \geq -x$ for all $x \in R$.

   Proof: By Trichotomy either $x > 0$ or $x = 0$ or $x < 0$. The result is trivial if $x = 0$; if $x > 0$ we have $|x| = x > 0 > -x$; if $x < 0$ we have $|x| = -x > 0 > x$. In every case, $|x| \geq$ both x and $-x$. □

The next observation is important because it allows us to interpret inequalities involving absolute values in terms of the geometry of the number line, and apply visual intuition to solve them.

5. If $r > 0$ then $|x| < r \Leftrightarrow -r < x < r$ (so x lies between the points $-r$ and r in the number line).

Proof ($\Leftarrow$): We know that $|x| = +x$ or $-x$. Multiplying each term in the given inequality $-r < x < r$ by (-1) we get $-r < -x < r$, therefore $|x| = \pm x$ is $< r$ in both cases.

Proof ($\Rightarrow$): By Trichotomy there are two cases

Case 1: $x \geq 0$. Then $r > |x| = x$, and $r > |x| \geq -x$ by (4.) Multiplying both sides of this last inequality by (-1) we obtain $-r < -(-x) = x$ too. Thus $-r < x < r$.

Case 2: $x < 0$. Then $-x > 0$. Since $|x| \geq \pm x$ by (4.) we get

$$\left.\begin{array}{r} r > |x| = -x \ \Rightarrow \ -r < -(-x) = x \\ r > |x| \geq x \ \Rightarrow \ x < r \end{array}\right\} \Rightarrow \ -r < x < r \quad \square$$

Other basic properties of absolute value are posted in the following exercise.

***2.2.2 Exercise.** Show that absolute value $|x|$ in a commutative ordered ring has the following properties.

(a) $|xy| = |x| \cdot |y|$

(b) $|x^2| = |x|^2 = x^2$

(c) Triangle Inequality: $|x \pm y| \leq |x| + |y|$, for all $x, y \in \mathbb{Z}$.

Hint: In (a) and (b) you might do casework based on the signs of x or y. In (c) it suffices to prove $|x+y| \leq |x| + |y|$; since $|-y| = |y|$ the other version follows upon replacing $y \mapsto -y$. For the "+" version, square both sides, do some algebra, and use Exercise 2.1.9. □

2.2.3 Lemma. *If $x \neq 0$ in $\mathbb{Z}$ then $|x| \geq 1$. Furthermore, $|x| = 1 \Leftrightarrow x = +1$ or $x = -1$.*

Proof: $\mathbb{Z}$ is a disjoint union $-\mathbb{N} \cup \{0\} \cup \mathbb{N}$; by our remarks in 2.1.5, the absolute value $|x|$ is in $\mathbb{N}$ if $x \neq 0$, and hence by 2.1.12 we must have $|x| \geq 1$. The second statement is obvious once you observe that $|x|$ is either $+x$ or $-x$. □

The **units**, or **invertible elements**, in $\mathbb{Z}$ are those $u \in \mathbb{Z}$ that have a multiplicative inverse: there exists some element $v \in \mathbb{Z}$ such that $uv = 1$. As an application of 2.2.3 we determine the units in $\mathbb{Z}$.

2.2.4 Lemma. The only units in $\mathbb{Z}$ are $+1$ and -1.

Proof: If $uv = 1$, then $|u| \cdot |v| = 1$ and neither absolute value can be zero. (by 2.1.2(#5) and the fact that $1 \neq 0$). By 2.1.12, we have $|u| \geq 1$ and $|v| \geq 1$.

If either absolute value were greater than 1, say $|u| > 1$, we would have $1 = |uv| = |u| \cdot |v| > 1 \cdot |v| = |v| \geq 1$. The net result would be that $1 > 1$, which is impossible. Conclusion: neither absolute value can exceed 1, and hence $|u| = |v| = 1$. Applying Lemma 2.2.3 we see that $u = \pm 1$. □

Divisibility in the system of integers. We begin with the official definition of "divisibility," keeping an eye on the exceptional role of the zero element.

2.2.5 Definition. If a and b are two integers we say that b **divides** a, often written as $b|a$, if there exists an $m \in \mathbb{Z}$ such that $a = mb$. Another way to put it: $b|a \Leftrightarrow a$ is a *multiple of* b. □

Here are some easy consequences of this definition.

2.2.6 Exercise. Verify the following facts from the definition.

1. $a|0$ for any $a \in \mathbb{Z}$.
2. 0 does not divide any non-zero element in $\mathbb{Z}$. (However, we do have $0|0$ according to our definition.)
3. $1|a$ and $-1|a$ for all $a \in \mathbb{Z}$. (Thus ± 1 are "trivial divisors" of every $a \in \mathbb{Z}$)
4. Divisibility is transitive: $a|b$ and $b|c \Longrightarrow a|c$ □

2.2.7 Lemma. *If $a, b \neq 0$ then $a|b$ and $b|a \Longrightarrow b = a$ or $b = -a$.*

Proof: If $a|b$ and $b|a$, there exist $c_1, c_2 \in \mathbb{Z}$ such that $a = c_1 b = c_1(c_2 a) = (c_1 c_2)a$. Since a is non-zero, by the cancellation law 2.1.7(b) we have $c_1 c_2 = 1$, and hence c_1 and c_2 are both units in $\mathbb{Z}$. By 2.2.4, $c_2 = \pm 1$ and therefore $b = c_2 a = \pm a$, as claimed. □

The next result is fundamental in all discussions of algebra and number theory.

2.2.8 Theorem (Euclidean Division Algorithm). *Let $a, b \in \mathbb{Z}$ with $b \neq 0$. Then there exist m, r such that*

$$a = mb + r \quad \textit{with} \quad m \in \mathbb{Z} \quad \textit{and} \quad 0 \leq |r| < |b|. \tag{5}$$

We can always arrange that $r \geq 0$, and then the pair (m, r) is unique.

Proof: If $a = 0$ we can take $m = r = 0$, so we may assume a is non-zero. To get $r \geq 0$, suppose we have $a = mb + r$ with $r < 0$. Then $-|b| < r < 0$; adding $|b|$ to both sides gives $0 < |b| + r < |b|$. Now rewrite (5) as

$$a = mb + r = (mb - |b|) + (|b| + r) = m'b + r' \quad \text{with } 0 \leq r' < |b|$$

Here $m'b = mb - |b| = mb \pm b$ is clearly just another multiple of b and the remainder is nonnegative, as desired.

As for uniqueness, suppose $r \geq 0$ and that we have two decompositions $a = m_1 b + r_1 = m_2 b + r_2$. We may label things so $r_2 \geq r_1$, and then $0 \leq r_2 - r_1 \leq r_2 < |b|$. Taking the difference of the two identities we get $r_2 - r_1 = (m_1 - m_2)b$; taking absolute values we get $0 \leq |m_1 - m_2| \cdot |b| = r_2 - r_1 < |b|$. The only way this can happen is to have $m_1 - m_2 = 0$ and $m_1 = m_2$, and then it follows immediately that $r_2 = r_1$.

In practice, division with remainder is most often performed for $a, b > 0$. Besides, if we prove (5) in this case the result for $a < 0$ follows upon multiplying both sides by (-1).

$$-a = (-m)b + (-r) \qquad \text{with } |-r| < |b|$$

and the result for $b < 0$ follows if we rewrite $mb = (-m)(-b)$

$$a = (-m)(-b) + r \qquad \text{with } |r| < |-b| = |b|$$

Assuming $a \geq 0$ and $b > 0$ we execute the following algorithmic steps to get (5).

Step 1. If $0 \leq a < b$, exit and report: $a = 0 \cdot b + a$, in which $m = 0, r = a$. Otherwise we have $a \geq b$ and go to the next step.

Step 2. Since $a \geq b$ we now test whether $b \leq a < 2b$. If so, exit and report

$a = 1 \cdot b + (a - b)$ in which $m = 1$ and $0 \leq r = a - b < 2b - b = b$. Otherwise, we have $a \geq 2b$ and we go to the next step.

Step 3. Since $a \geq 2b$ we now test whether $a < 3b$. If so, exit and report $a = 2 \cdot b + (a - 2b)$ in which $m = 2, 0 \leq r = a - 2b < b$. Otherwise, go to the next step.

$$\vdots$$

We have indicated an inductive process which could easily be converted into computer code. The only unresolved issue is whether this process terminates in some finite number of steps.

To see why it does, consider what would happen if it did not. Then for every $n \in \mathbb{N}$ our number a would fail the test $a \leq nb$. That means $a > nb$ for all $n \in \mathbb{N}$, and since $b \geq 1$ we would be forced to conclude that $a > nb \geq n$ for all counting numbers $n \in \mathbb{N}$! That is impossible, because $a \in \mathbb{N}$ and we obtain the contradiction $a + 1 > a > a + 1$. □

This proof outlined an algorithm for finding m and r, at least when a and b are positive: simply keep subtracting copies of b from a until $0 \leq a - mb < b$. Further analysis yields upper bounds on the number of steps required to execute the algorithm. (It turns out that this is quite a fast algorithm, and is essentially what goes on inside your calculator.)

If you have a calculator, here's a fast way do Euclidean Division:

1. Compute the decimal version of
$$\frac{a}{b} = b_n \ldots b_0 \,.\, a_1 a_2 \ldots = (\textit{integer part}) + (\textit{fractional part})$$
2. Store (*integer part*) and (*fractional part*).
3. Then $r = |b| \cdot (\textit{fractional part})$ is an integer such that $0 \leq r < |b|$ and we have $a = mb + r$.

Your calculated answer for r may exhibit trailing "nines" due to calculator round-off error; round up to get r as an integer.

2.2.9 Exercise. Use this calculator routine to write $a = mb + r$ with $0 \leq r < |b|$ when

(a) $a = 11,473$ and $b = 598$ (b) $a = 11,473$ and $b = -598$

***2.2.10 Exercise.** Taking $a = 5, b = 7$ show that m and r in $7 = 5m + r$ are not uniquely determined if we only require $0 \leq |r| < |b|$. □

2.2.11 Definition (Greatest Common Divisor). *A* **greatest common divisor** *of nonzero elements $a, b \in \mathbb{Z}$, denoted by (a, b) or $\gcd(a, b)$, is an element $c \in \mathbb{Z}$ such that*

1. *$c|a$ and $c|b$*
2. *$c > 0$*
3. *If c' is any other element of $\mathbb{Z}$ satisfying (1.) and (2.) then c' divides c.*

A greatest common divisor is denoted by $\gcd(a, b)$, or sometimes just (a, b).

The following lemma demonstrates existence of $\gcd(a, b)$ for any nonzero integers a, b, and its proof provides a handy geometric interpretation of the *gcd*.

2.2.12 Lemma. *For any pair of nonzero integers there exists a unique greatest common divisor $\gcd(a, b)$, which has the form $\gcd(a, b) = ra + sb$ for suitably chosen $r, s \in \mathbb{Z}$.*

Proof: Uniqueness. Suppose there were two elements c and c' with the properties listed in 2.2.11. By Property 3, we would have $c|c'$ and $c'|c$. By Lemma 2.2.7 we have $c' = \pm c$. Since a *gcd* must be positive we get $c = c' > 0$.

Existence: In $\mathbb{Z}$ we construct the "additive lattice" $\Lambda = \Lambda(a, b)$ obtained by taking all "integer linear combinations" of the elements a and b.

$$\Lambda = \mathbb{Z}a + \mathbb{Z}b = \{ma + nb : m, n \in \mathbb{Z}\} \tag{6}$$

Obviously Λ contains $a = 1a + 0b$ and $b = 0a + 1b$. In fact there exist *positive* elements in Λ, so $\Lambda \cap \mathbb{N} \neq \emptyset$, because both $\pm a$ lie in Λ. By the Minimum Principle (Theorem 2.1.11) there exists a smallest element in $\Lambda \cap \mathbb{N}$, which we denote by c. By definition of Λ there are integers $m_0, n_0 \in \mathbb{Z}$ such that $c = m_0 a + n_0 b$.

We claim that c is a *gcd* of a and b. Statement (2.) in the definition of *gcd* is obviously true. We verify (1.) by applying the Euclidean Division Algorithm to prove a more general statement, namely

For any $ma + nb \in \Lambda$ the element c divides $ma + nb$

In fact, by the Division Algorithm 2.2.8 there exist k, r such that

$$\begin{aligned} ma + nb &= kc + r, \qquad \text{with } 0 \leq r < c \\ r &= ma + nb - kc \\ &= ma + nb - k(m_0 a + n_0 b) = (m - km_0)a + (n - kn_0)b \end{aligned}$$

This shows that $r \in \Lambda$. But $0 \leq r < c$, so by minimality of c we must have $r = 0$.

Finally, suppose $c' \in \mathbb{Z}$ divides both a and b. Then then $c'|x$ for any $x \in \Lambda$, and in particular $c'|c$. Thus c is a *gcd* of a and b. The proof is complete. □

2.2.13 Corollary. *If $a, b \in \mathbb{Z}$ with $a \neq 0$, then there exist $r, s \in \mathbb{Z}$ such that* $\gcd(a, b) = ra + sb$. *The greatest common divisor c is the smallest positive element in the lattice Λ defined in* (6) *above. Furthermore every element of Λ is a multiple of c, so that $\Lambda = \mathbb{Z} \cdot c$.*

PROOF: Since $c \in \Lambda$ it is obvious that $\mathbb{Z} \cdot c \subseteq \Lambda$. Conversely if $x \in \Lambda$ Euclidean division with remainder allows us to write $x = mc + r$ with $0 \leq r < c$, and then $r = x - mc \in \Lambda$ is smaller than the smallest positive element c in Λ. Therefore we must have $r = 0$, x is a multiple of c, and $\Lambda \subseteq \mathbb{Z} \cdot c$. □

***2.2.14 Exercise.** Let a, b be nonzero integers. Prove that their greatest common divisor has the following properties

(a) $\gcd(a, b) = \gcd(b, a)$

(b) $\gcd(a + kb, b) = \gcd(a, b)$ for any $k \in \mathbb{Z}$.

(c) If $a, b \neq 0$ and a divides b, show that $\gcd(a, b) = a$.

Hint: Use the fact that $c = \gcd(a, b)$ is the smallest positive element in $\Lambda = \mathbb{Z}a + \mathbb{Z}b$. Compare the sets $\Lambda = \mathbb{Z}a + \mathbb{Z}b$ and $\Lambda' = \mathbb{Z}(a + kb) + \mathbb{Z}b$, giving separate arguments to show that $\Lambda' \subseteq \Lambda$ and $\Lambda \subseteq \Lambda'$ to conclude that these sets are the same. □

The properties (b) and (c) in the last exercise are the basis of a very fast algorithm for determining the *gcd* of two integers $a, b > 0$.

> **The GCD Algorithm (sketch).** We may label things so $b \geq a$. If $a = b$ there's nothing to do, so suppose $0 < a < b$. We begin the recursive process that yields the *gcd* by labeling $b_0 = b > a_0 = a$; we produce the next pair of numbers $b_1 > a_1$ in the following way. As in the Euclidean division algorithm, subtract multiples of a from b until we get a remainder $0 \leq r = b - ka < a$. If $r = 0$ then b divides a so we have $\gcd(a, b) = a$ by (c) above, and we're done. Otherwise we get $\gcd(a, b) = \gcd(a, b - ka) = \gcd(a, r) = \gcd(r, a)$ with $0 < r < a$, and we are back in the situation we started with, except that the new integers are $b_1 = a_0 = a$ and $a_1 = r$ with $b_1 > a_1$.
>
> Both new numbers a_1 and b_1 are less than the largest number b in the original pair b_0, a_0, while $\gcd(a_0, b_0) = \gcd(a_1, b_1)$, so we have in effect decreased the size of the larger number without changing the gcd. Such reduction can only continue for

a finite number of steps. It terminates when we get a remainder of zero, in which event $a_m | b_m$ and

$$\gcd(a, b) = \gcd(a_0, b_0) = \gcd(a_1, b_1) = \ldots \gcd(a_m, b_m) = a_m \text{ (when } a_m | b_m).$$

2.2.15 Example. Use the GCD Algorithm to compute the greatest common divisor of 48 and 347 using the GCD algorithm.

Discussion: Labeling $a_0 = 48 < b_0 = 347$ we have

$$\begin{array}{rclclcl l}
\gcd(48, 347) & = & \gcd(48, 11) & = & \gcd(11, 48) & & & \text{because } 11 = 347 - 7(48) \\
& = & \gcd(11, 4) & = & \gcd(4, 11) & & & \text{because } 4 = 48 - 4(11) \\
& = & \gcd(4, 3) & = & \gcd(3, 4) & & & \text{because } 3 = 11 - 2(4) \\
& = & \gcd(3, 1) & = & \gcd(1, 3) & & & \text{because } 1 = 4 - 1(3) \\
& = & 1 & & & & &
\end{array}$$

Therefore $\gcd(48, 347) = 1$. □

It is possible to extract from the GCD algorithm (essentially by working it backward) a fast algorithm that produces integers r, s such that

$$ra + sb = \gcd(a, b)$$

The r, s need not be positive even if $a, b > 0$. For some purposes it is just as important to know the multipliers r, s as it is to determine the *gcd*. In fact, suppose that integers a and $n > 1$ are *relatively prime* in the sense that $\gcd(a, n) = 1$, so that $ra + sn = 1$ for suitably chosen $r, s \in \mathbb{Z}$. Then in $\mathbb{Z}_n$ we have found a multiplicative inverse for $[a]$. In fact, we have $sn \equiv 0 \pmod{n}$ so that

$$1 = ra + sn \equiv ra + 0 = ra \qquad \pmod{n}$$

Thus $[r] \cdot [a] = [1]$ and $[a]^{-1} = [r]$ (also $[r]^{-1} = [a]$). Finding multiplicative inverses in $\mathbb{Z}_n$ by trial and error, if they actually exist, is quite tedious; the systematic approach outlined above is much more efficient. The necessary modifications of the GCD Algorithm are given below.

Extended GCD Algorithm. The idea that allows us to find suitable r, s is that at the last step in the GCD Algorithm it is easy to write $c = \gcd(a_m, b_m)$ in the form $r_m a_m + s_m b_m = c$ because we are dealing with very small numbers. For instance in the last example $1 = \gcd(3, 1)$ can be seen by inspection to equal

$$1 = 1\boxed{3} + (-2)\boxed{1} \quad \text{with} \quad r = 1, s = -2.$$

Next we use the result $1 = 4 - 1(3)$ from the previous step in GCD to rewrite this, replacing the *smaller* of the two numbers $\boxed{**}$. This yields the identity

$$\boxed{3} + (-2)\big[\,4 - 1(3)\,\big] = 3\boxed{3} + (-2)\boxed{4} = 1$$

Continuing, we again replace the smaller number $\boxed{3}$ in this identity, this time using the next-to-last identity $3 = 11 - 2(4)$ in the GCD process. We get

$$3\big[\,11 - 2(4)\,\big] + (-2)\boxed{4} = 3\boxed{11} - 8\boxed{4} = 1$$

In the next step we replace $4 = 48 - 4(11)$ to get

$$3\boxed{11} - 8\big[\,48 - 4(11)\,\big] = 35\boxed{11} - 8\boxed{48} = 1$$

Finally, we replace $11 = 347 - 7(48)$ to get

$$35\,\big[\,347 - 7(48)\,\big] - 8\boxed{48} = 35\boxed{347} - 253\boxed{48} = 1$$

This is the answer to our prayers: the *gcd* $= 1$ has been expressed in the form $347r + 48s = 1$, with very little additional effort once the forward GCD Algorithm has been executed to find $\gcd(347, 48)$. (You must however retain the computational details recorded in the right-hand column of Example 2.2.15.) □

***2.2.16 Exercise.** Use the algorithm outlined above to find $\gcd(a, b)$ for the following pairs of integers

(a) $a = 5, b = 85$ (b) $a = 296, b = 1317$ (c) $a = 955, b = 11422$

By trial and error (using a calculator) see if you can find integers r, s such that $ra+sb = \gcd(a, b)$ in each case.
Note: r, s need not be positive. □

***2.2.17 Exercise.** Use the GCD algorithm to verify that $\gcd(28, 15) = 1$. Then find some pair of integers $r_0, s_0 \in \mathbb{Z}$ such that $15r_0 + 28s_0 = 1$. Finally, determine *all* pairs (r, s) such that $15r + 28s = 1$.
Hint: You could use trial and error or the Extended GCD Algorithm to find r_0, s_0, since we are dealing with small values of a, b. □

***2.2.18 Exercise.** Use the Extended GCD Algorithm to find r, s such that $ra + sb = \gcd(a, b)$ for each of the pairs a, b listed in Exercise 2.2.11. □

2.2.19 Exercise. Generalize the definition of *gcd* to define $\gcd(a_1, \ldots, a_r)$ where the a_i are nonzero. Make the obvious changes in the definition of $\gcd(a, b)$ and then

(a) Prove $c = \gcd(a_1, \ldots, a_r)$ exists by considering the set of integer linear combinations

$$\Lambda = \mathbb{Z}a_1 + \ldots + \mathbb{Z}a_r = \left\{ \sum_{i=1}^{r} k_i a_i : k_i \in \mathbb{Z} \right\}$$

Show that $\Lambda \cap \mathbb{N} \neq \emptyset$ and verify that the smallest element $c \in \Lambda \cap \mathbb{N}$ (which exists by the Minimum Principle) has the properties required of $\gcd(a_1, \ldots, a_r)$.

(b) Show that $\Lambda = \mathbb{Z} \cdot c =$ all integer multiples of $\gcd(a_1, \ldots, a_r)$

We say that $a_1, \ldots, a_r$ are **jointly relatively prime** if $\gcd(a_1, \ldots, a_r) = 1$. □

As we will soon see, the a_i are jointly relatively prime $\Leftrightarrow$ no prime $p > 1$ is a common divisor of all the integers a_i.
Note: The proof is a simple variant of the previous discussion in these *Notes*. □

***2.2.20 Exercise.** Integers $a_1, \ldots, a_n > 0$ are said to be **pairwise relatively prime** if $\gcd(a_i, a_j) = 1$ for all $i \neq j$. Intuitively, this means no pair a_i, a_j has a prime divisor in common.

(a) Which is the stronger condition: *pairwise relatively prime* or *jointly relatively prime*? Explain.

(b) Using your prior knowledge about primes in the integers give an example of three integers a_1, a_2, a_3 that are jointly relatively prime but not pairwise relatively prime.

***2.2.21 Exercise.** We will soon show that $[k] \in \mathbb{Z}_n$ has a multiplicative inverse $\Leftrightarrow \gcd(k, n) = 1$. For each of the following pairs, check that *gcd* $= 1$ and determine the multiplicative inverse $[k]^{-1}$

(a) $k = 6, n = 45$ (b) $k = 48, n = 1127$ (c) $k = 296, n = 1317$

***2.2.22 Exercise.** Find the *gcd* of 48 and 256. Show that $[48]$ cannot have a multiplicative inverse in $\mathbb{Z}_{256}$ by showing that it is a zero divisor: there exists an ℓ such that $[48] \cdot [\ell] = [0]$.

Hint: Write $gcd = 48r + 256s$ and use this to show [48] is a zero divisor. □

2.3 Primes and Prime Factorization in $\mathbb{Z}$.

Recall that a *unit* in $\mathbb{Z}$ is an element that has a multiplicative inverse: there exists some $v \in \mathbb{Z}$ such that $uv = 1$. Thus the units are the *divisors of the identity element* 1. In 2.2.4 we showed that 1 and -1 are the only units in $\mathbb{Z}$. Later on we will examine more general structures $(R, +, \cdot\,)$, and employ a similar definition of the units in R. Then the discussion of units becomes much more interesting.

2.3.1 Definition. *A* **prime** *is an element p in $\mathbb{Z}$ that cannot be written as a product $p = ab$ both of whose factors are non-units.*

If an integer $b \neq 0$ can be written as $b = cd$ with both c and d nonunits, we refer to this as a **nontrivial factorization** of b. From these definitions it follows that p is a prime $\Leftrightarrow$ $-p$ is prime; however, our main concern will be with primes $p \geq 1$. By this definition -1 and 1 are primes (why?). Sometimes it is convenient to regard these as the "trivial primes" and exclude them.

2.3.2 Definition. Explain why an integer $n \neq 0$ is prime $\Leftrightarrow$ $-n$ is a prime. □

2.3.3 Definition. *If a, b are nonzero elements in $\mathbb{Z}$ we say they are* **relatively prime** *if* $\gcd(a, b) = 1$. *That means: there exist $r, s \in \mathbb{Z}$ such that $ra + sb = 1$.*

Equivalently, a, b have no common divisor $c > 1$. Verifying that the usual list $1, 2, 3, 5, 7, 11, \ldots$ consists of primes it not so straightforward if you try to do it directly from the axioms.

2.3.4 Exercise. If $n > 1$ has a nontrivial factorization $n = ab$ (a and b not equal to ± 1) then $|a|, |b| > 1$ and by 2.1.14 we must have $|a| \geq 2$ and $|b| \geq 2$. Use this observation to prove that

(a) $2 = 1 + 1$ is a prime

(b) $3 = 2 + 1 = 1 + 1 + 1$ is a prime

(c) $4 = 3 + 1 = 1 + 1 + 1 + 1$ is *not* a prime

Hints: Use basic facts 2.1.6 about inequalities and basic facts 2.2.2 about absolute values. To deal with (c) you will also need the fact that $2 \cdot 2 = 4$. How do you prove that from the axioms? (What is the definition of "4"?)
Note: We're not quite ready to tackle more difficult questions – e.g. are 16 or 17 prime? □

2.3.5 Exercise. For $a, b \neq 0$ in $\mathbb{Z}$, verify that

If $a' = u_1 a$ and $b' = u_2 b$ with u_1, u_2 units in $\mathbb{Z}$, then $\gcd(a, b) = \gcd(a', b')$.

Thus $\gcd(a, b)$ is unaffected if the entries a and b are multiplied by arbitrary units. □

2.3.6 Exercise. Let $p > 1$ be a prime. Show that its only divisors are $\pm p$ and ± 1. □

We now take up the question of prime factorization. You have all been told that every integer $n > 1$ has a unique factorization into primes $p > 1$. The big question is: How do we deduce that piece of folklore from the axioms? The proofs go back to Euclid. The first step is to prove that there exists at least one such factorization

2.3.7 Theorem (Prime Factorization I). *Every integer $n > 1$ can can be written as $n = p_1 p_2 \ldots p_r$ where each factor is a prime $p_i > 1$.*

Some preliminary comments are in order. We will argue by induction starting with $n = 2$ rather than $n = 1$. The initial case is easy: $n = 2$ is already a prime (see 2.3.4). But the induction step involves a slight twist. You might expect to procede using Induction to verify the statements

$P(n)$ = (The integer n has a factorization into nontrivial positive primes) for $n \geq 2$, but it turns out to be more efficient to prove that the stronger statements

$$Q(n) = (\text{Every integer } k \text{ with } 2 \leq k \leq n \text{ has a factorization into nontrivial positive primes})$$

are true. Obviously $Q(n) \Rightarrow P(n)$ for all n, so it is certainly alright to make this switch. The question is, why deal with the more complicated statements $Q(n)$ when our ultimate interest is in $P(n)$? The answer: Mathematics is an art rooted in intuition, experience, and sometimes plain animal cunning. It is easy to check a proof to see if it is valid; it can be hard to create one from scratch, without any idea what your first move should be. The switch from $P(n)$ to $Q(n)$ is an example of the art – an unexpected move that actually helps you get to your goal. Watch how it works below.

First we prove a preliminary lemma which will sound familiar, and is the basis of the induction proof that comes next.

2.3.8 Lemma. *Suppose $n > 1$ has a nontrivial factorization $n = ab$ with $a, b \in \mathbb{Z}$ (neither factor a unit). Then the absolute values of the factors must both be smaller than n:*

$$1 < |a| < n \qquad \textit{and} \qquad 1 < |b| < n$$

Proof: We can't have $|a| = 1$: by 2.2.4 that would mean $a = \pm 1$, which is excluded. Likewise for b. Since the absolute values lie in $\mathbb{N}$, it follows from 2.1.11 that $|a| > 1$ and $|b| > 1$. Then we see that $n = |n| = |ab| = |a| \cdot |b| > 1 \cdot |b| = |b|$, and similarly $n > |a|$. □

Proof of 2.3.7: As for the statements $Q(n)$, obviously $Q(2)$ is true. Next suppose $Q(n)$ is true for $n = n_0 \geq 2$; we want to show that $Q(n_0 + 1)$ is true. Let k be any integer $2 \leq k \leq n_0 + 1$. If $k \leq n_0$ it has a prime factorization because $Q(n_0)$ holds. If $k = n_0 + 1$ itself is a prime, $Q(n_0 + 1)$ is automatically true. [Why? Because in this situation $Q(k)$ is assumed true for all $k \leq n_0$, and it is also true for $k = n_0 + 1$. By 2.1.14 no integer can lie strictly between n_0 and $n_0 + 1$, so the statement is true for all $k \leq n_0 + 1$, verifying $Q(n_0 + 1)$.]

In the remaining case $k = n_0 + 1$ is not prime so there must be a nontrivial factorization $n_0 + 1 = ab$. Since $ab = |a| \cdot |b|$ we can assume $a, b > 0$, and then $a, b > 1$ since neither factor is a unit.

Applying 2.3.7 we get $1 < a < n_0 + 1$ and $1 < b < n_0 + 1$. By 2.1.15 no element in $\mathbb{N}$ can lie between n_0 and $n_0 + 1$, so we must actually have $a, b \leq n_0$. Therefore we can apply the (valid) statement $Q(n_0)$ to both factors, writing

$$a = p_1 \ldots p_r \qquad b = q_1 \ldots q_s \quad ,$$

where $p_i, q_j > 1$ are (not necessarily distinct) primes in $\mathbb{N}$. Hence $n_0 + 1$ is also a product of nontrivial positive primes: $n_0 + 1 = p_1 \ldots p_r q_1 \ldots q_s$, and $Q(n_0 + 1)$ is verified. That completes the inductive step of our proof. By the Induction Axiom, the statements $Q(n)$, and $P(n)$, must be true for all integers $n \geq 2$. □

Question: To see if you really understand the argument given above, can you explain why the discussion gets into trouble if you try to make it work using $P(n)$ instead of $Q(n)$?

2.3.9 Corollary (Euclid). *There exist infinitely many prime numbers $p > 1$ in $\mathbb{N}$.*

Proof: The proof is a classic example of "Argument by Contradiction," which works as follows. Any proposition P is either true or false. Suppose we assume P is *not true* (i.e. that $\neg P$ is true), and by valid reasoning from this premise arrive at an absurdity such as "$1 \neq 1$" or a conclusion such as "$0 = 1$" that is in conflict with one of the Axioms governing $\mathbb{Z}$. Then the only viable conclusion is that P *must be true.*

So, suppose our claim is wrong and there *are* only finitely many primes. Then the set of primes in $\mathbb{N}$ is bijectively equivalent to an interval $[1, n]$ and the primes can be listed as $p_1, p_2, \cdots, p_n$ for some $n \in \mathbb{N}$. Indeed, we may label our primes in increasing order so that

$p_i < p_{i+1}$, and then our list will read: $p_1 = 2$, $p_2 = 3$, $p_3 = 5, p_4 = 7, \ldots$ (recall Exercise 2.3.3). Now consider the integers

$$\begin{aligned} a &= p_1 p_2 \cdots p_n \qquad \text{(the product of all primes)} \\ q &= 1 + a \;=\; 1 + p_1 p_2 \cdots p_n \qquad \text{(the successor of } a\text{)} \end{aligned}$$

Finiteness of the set of primes is used here; the product defining a only makes sense if it has finitely many terms.

By the Factorization Theorem 2.3.7, there exist non-negative integers $s_i > 0$ such that

$$q = \prod_{i=1}^{n} p_i^{s_i} \qquad (p_i \text{ distinct primes} \geq 1)$$

Since $q > 1$ there must be at least one index $i > 1$ such that $s_i > 0$. In particular, $p_i | q$. Now observe that

$$\begin{aligned} &p_i \text{ divides } a = p_1 \cdots p_n \qquad \text{(from definition of } a\text{)} \\ &p_i \text{ divides } q = 1 + a \qquad \text{(by definition of the index } i\text{)} \end{aligned}$$

Hence p_i divides $1 = (1 + a) - a$ and there is some $b \in \mathbb{Z}$ such that $1 = b \cdot p_i$. That forces both b and p_i to be units in $\mathbb{Z}$, so that $p_i = \pm 1$ (by 2.2.4), which is impossible since $p_i > 1$. Conclusion: our original statement asserting the existence of infinitely many primes $p > 1$ must be true. □

Uniqueness of the prime factorization. We now show that the factorization described in 2.3.7 is unique. To prove this we need a few additional facts about $\gcd(a, b)$ when $a, b \neq 0$.

2.3.10 Lemma.
If $p > 1$ is a prime and n is a nonzero integer, then either $p|n$ or $\gcd(p, n) = 1$.

Proof: Suppose $c = \gcd(p, n) \neq 1$. Then $c > 1$ and c divides p, which means there exists some b such that $p = bc$. By the definition of the primes, at least one of the factors b and c must be a unit in $\mathbb{Z}$, hence equal to ± 1. It cannot be c because $c > 1$. Thus $b = \pm 1$ and $p = |p| = |bc| = |c| = c$. Then $p|n$ because $c|n$. □

2.3.11 Lemma. *Let a, b, c be non-zero integers. If $a|bc$ and $\gcd(a, b) = 1$ then $a|c$.*

Proof: Since $\gcd(a, b) = 1$ there exists integers r and s such that $1 = ra + sb$. Therefore $c = rac + sbc$. Furthermore $a|bc$, so that $a|sbc$. It follows that a divides $c = rac + sbc$. □

The following corollary is the key fact needed to prove uniqueness.

2.3.12 Corollary. *If a prime $p > 1$ divides a product $b_1 \cdots b_k$ of nonzero integers, then there exists an index i such that $p|b_i$.*

Proof: We argue by induction on the number k of factors in the product. If $k = 1$, we can simply pick $i = k = 1$. Next suppose our claim is true for an integer $k \geq 1$; we must prove it true for $k + 1$. In this induction step we are assuming that $p|(b_1 \cdots b_k)b_{k+1}$. By Lemma 2.3.10, either $p|b_{k+1}$, in which case our claim is proved, or $\gcd(p, b_{k+1}) = 1$. In the latter case we apply Lemma 2.3.11 to get $p|(b_1 \cdots b_k)$. By our induction hypothesis, there is some factor b_i, $i \leq k$, such that $p|b_i$. Thus our claim is true in this case too. The proof is complete. □

2.3.13 Theorem (Unique Factorization II). *Any integer $n > 1$ has a unique factorization as a product of nontrivial positive primes $p > 1$. That means,*

$$n = \prod_{i=1}^{m} p_i^{r_i} \qquad \textit{where} \begin{cases} p_1, p_2, \ldots, p_m > 1 \textit{ are distinct primes} \\ r_i > 0 \textit{ for all } 1 \leq i \leq m \end{cases} \tag{7}$$

The primes p_i appearing in (7) are unique as are their "multiplicities," the exponents r_i.

Proof: Suppose we have two factorizations

$$n = \prod_{i=1}^{r} p_i^{m_i} = \prod_{j=1}^{s} q_j^{n_j} \qquad p_i, q_j > 1 \text{ primes, } m_i, n_j \geq 1$$

First note that the sets of distinct primes $\{p_i\}, \{q_j\}$ appearing in the factorization of n must coincide. In fact if we had, say, $p_i \notin \{q_j\}$ for some index i, then $p_i|n$ because p_i divides the left-hand product, but then p_i would divide the right-hand product, which is impossible by 2.3.12. (By 2.3.6 the only divisors of the prime q_j are $\pm q_j$ and ± 1, so $p_i = q_j$ for some j.)

Once we know that $r = s$ we can rewrite the products as

$$n = \prod_{i=1}^{r} p_i^{m_i} = \prod_{i=1}^{r} p_i^{n_i} \qquad p_i > 1 \text{ distinct primes, } m_i, n_j \geq 1$$

The factorizations can differ only in their multiplicities. If $m_i = n_i$ for all i, we're done. Otherwise there is an index $1 \leq k \leq r$ such that $m_i = n_i$ for $i < k$ and $m_k \neq n_k$. (We may assume $m_k > n_k$ by relabeling.) Both products now have a common factor $\prod_{i=1}^{k-1} p_i^{m_i}$, which we may cancel by 2.1.7 to get

$$\prod_{i=k}^{r} p_i^{m_i} = \prod_{i=k}^{r} p_i^{n_i} \qquad \text{with } m_k > n_k \geq 1$$

There is still a common factor $p_k^{n_k}$ on both sides. Cancelling this we get

$$p_k^{m_k - n_k} \cdot \prod_{i=k+1}^{r} p_i^{m_i} = \prod_{i=k+1}^{r} p_i^{n_i} \qquad \text{with } m_k > n_k \geq 1$$

Once again we are in conflict with 2.3.12: p_k appears in the left-hand product but not in the product on the right. We arrived at this contradiction by assuming that the multiplicities did not match, so we have proved the theorem. □

Using 2.3.13 and the preceding lemmas we can compute $\gcd(a, b)$, or decide when $\gcd(a, b) = 1$, if we know the prime factorizations of a and b. The necessary results are developed in the next exercises.

∗**2.3.14 Exercise.** Let $a > 1$, $b > 1$ be integers and let

$$a = \prod_{i=1}^{m} p_i^{r_i} \qquad b = \prod_{j=1}^{k} q_j^{s_j} \tag{8}$$

be their unique factorizations into nontrivial positive primes. Prove that $\gcd(a, b) = 1 \Leftrightarrow$ these factorizations have no primes in common, so that $\{p_1, \ldots, p_m\} \cap \{q_1, \ldots, q_k\} = \emptyset$ – i.e. if no single prime $p > 1$ divides both a and b. □
Hint: As stated, our claim has the form $P \Leftrightarrow Q$. It might be easier to prove $\neg P \Leftrightarrow \neg Q$o □.

∗**2.3.15 Exercise.** Let $p > 1$ be a prime and let $a \neq 0$. Prove that $p|a \Leftrightarrow p^2|a^2$. In particular, a nonzero integer a is even (or odd) $\Leftrightarrow a^2$ is even (or odd). □

∗**2.3.16 Example.** In $\mathbb{R}$ the number $\sqrt{2}$ is irrational

Discussion: Suppose we could represent $\sqrt{2} = p/q$ where p, q are integers such that $q \neq 0$. By cancelling common prime factors and invoking 2.3.14, we can assume that p and q have no prime factors in common, with $\gcd(p, q) = 1$. Squaring and then multiplying both sides by q^2 we get $2q^2 = p^2$. This immediately implies that p^2 is even, and hence by p must be even. That means $p = 2p'$ for some p' so $2q^2 = (2p')^2 = 4(p')^2$. Both sides are divisible by 2, yielding $q^2 = 2(p')^2$. This implies that q^2, and hence also q, is even. But if p and q are both even they must have "2" as a common prime factor. This is impossible because the representatives p and q were chosen at the outset to be relatively prime. □

$*$**2.3.17 Exercise.** Adapt the discussion of 2.3.16 to prove that $\sqrt{3}$ and $\sqrt{6}$ are irrational.
Hint: If there exist $a, b > 0$ such that $\sqrt{3} = a/b$, so that $a^2/b^2 = 3$, we can cancel common factors and assume that $\gcd(a, b) = 1$. Then $3a^2 = b^2 \Rightarrow 3|b^2$, etc. □

2.3.18 Exercise. If $n \geq 0$ prove that $\sqrt{n}$ is irrational unless n is a "perfect square" $n = a^2$ for some $a \in \mathbb{Z}$. □

***2.3.19 Exercise.** Let $a, b > 1$ be integers and suppose there are some common primes in the unique factorizations (8). If the common primes are $u_1, \ldots, u_\ell$ let's relabel terms in (8) as follows

$$a = u_1^{r_1} \cdots u_\ell^{r_\ell} \cdot p_{\ell+1}^{r_{\ell+1}} \cdots p_m^{r_m} \qquad b = u_1^{s_1} \cdots u_\ell^{s_\ell} \cdot q_{\ell+1}^{s_{\ell+1}} \cdots q_k^{s_k} \quad (\text{with } r_i, s_i > 0)$$

where $\{p_{\ell+1}, \ldots, p_m\} \cap \{q_{\ell+1}, \ldots, q_k\} = \emptyset$. (Either of the sets $\{p_i\}, \{q_j\}$ might be empty.) Prove that

$$(9) \qquad \gcd(a, b) = u_1^{c_1} \cdots u_\ell^{c_\ell} \qquad \text{where} \qquad c_i = \min\{r_i, s_i\} \text{ for each index } 1 \leq i \leq \ell \quad \square$$

Note: The *gcd* is 1 if there are no prime divisors in common. □

***2.3.20 Exercise.** If $a, b \neq 0$ are *not* relatively prime, prove that

$$a' = \frac{a}{\gcd(a, b)} \qquad \text{and} \qquad b' = \frac{b}{\gcd(a, b)}$$

are relatively prime. □

The following observation is useful in finding prime factorizations because it limits the range of integers that must be examined to find the smallest prime divisor of a non-prime integer.

***2.3.21 Exercise.** If $n = q_1 \cdots q_r$ with each $q_i > 1$ prime (repeats allowed) and $r \geq 2$ (so n is not already prime), show that there is some index i such that $p_i \leq \sqrt{n}$. □

***2.3.22 Exercise.** Which of the following integers are prime? If not prime, find the prime factorization

(a) $n = 17$ (b) $n = 517$ (c) $n = 518$ (d) $n = 54$ (e) $n = 1159$ □

2.3.23 Exercise. List all primes such that $1 < p < 200$. □

***2.3.24 Exercise.** If $p > 1$ is a prime and if $n \neq 0$ prove that $\gcd(p, n) \neq 1 \Leftrightarrow p$ divides n.
Hint: Recall 2.3.5. □

2.3.25 Exercise. If $a, b, c \neq 0$ is it true that $\gcd(a, b, c) = \gcd(a, \gcd(b, c))$? Prove or provide a counterexample.
Note: You could explore this using unique factorization (as in 2.3.19), or using the basic definitions of *gcd* in terms of divisibility (as in 2.2.11 or 2.2.17). □

2.3.26 Exercise. In terms of the description 2.3.17 of *gcd* in terms of prime factorizations: explain the difference between a collection of integers $a_1, \ldots, a_n \neq 0$ being "jointly relatively prime," so $\gcd(a_1, \ldots, a_n) = 1$, and being "pairwise relatively prime," so $\gcd(a_i, a_j) = 1$ for $i \neq j$. □

2.3.27 Exercise. The **least common multiple** $\operatorname{lcm}(a_1, \ldots, a_n)$ of integers $a_i \neq 0$ is the smallest positive integer d (in the sense of divisibility) such that $a_i|d$ for all i.

(a) Give a precise version of this definition and prove that the *lcm* is unique if it exists.

(b) If $p_1, \ldots, p_r$ are distinct primes and positive integers $a, b > 1$ have prime decompositions $a = \prod_{i=1}^{r} p_i^{m_i}$, $b = \prod_{i=1}^{r} p_i^{n_i}$ (with $m_i, n_i \geq 0$, allowing some multiplicities to be zero), prove that $\mathrm{lcm}(a, b)$ exists and is given by

$$\mathrm{lcm}(a, b) = \prod_{i=1}^{r} p_i^{s_i} \qquad \text{where } s_i = \max\{m_i, n_i\} \text{ for each } i$$

What happens when $a = 1$ or $b = 1$?

(c) If $a, b, c > 1$ how do you describe $\mathrm{lcm}(a, b, c)$ in terms of the prime factorizations?

(d) If $a, b, c > 1$ is it true that $\mathrm{lcm}(a, b, c) = \mathrm{lcm}(a, \mathrm{lcm}(b, c))$? Prove or provide a counterexample.

Note: Recall that a result similar to (b) holds for $\gcd(a, b)$. □

2.4 Modular Arithmetic in $\mathbb{Z}_n$ Revisited.

In Section 1.5 of Chapter 1 we introduced the concept of an *equivalence*, or RST *relation* $x \sim y$ in a set X, and the associated *quotient space* X/R consisting of the equivalence classes

$$[a] = \{x \in X : x \sim a\}$$

determined by the relation. We then explained how to create a new algebraic systems $(\mathbb{Z}_n, +, \cdot\,)$ for each integer $n \in \mathbb{N}$ by imposing the "congruence relation" $a \equiv b \pmod n$ on the system of integers $(\mathbb{Z}, +, \cdot\,)$. We recall that definition:

2.4.1 Definition (Congruence mod n). *Fix an integer $n > 1$ and define the following* RST *relation in $X = \mathbb{Z}$:*

$$\begin{aligned} a \equiv b \ (mod\ n) \quad &\Leftrightarrow \quad b - a \text{ is a multiple of } n \\ &\Leftrightarrow \quad b = a + nk \text{ for some } k \in \mathbb{Z} \\ &\Leftrightarrow \quad b \in a + n\mathbb{Z} = \{a + nk : k \in \mathbb{Z}\} \\ &\Leftrightarrow \quad b + n\mathbb{Z} = a + n\mathbb{Z} \end{aligned} \tag{10}$$

It is easily seen that this defines a relation that is reflexive, symmetric, and transitive. In plain English, the relation $a \equiv b$ (mod n) is read as: "a is **congruent to** *b* **modulo the integer** *n," and for this reason the equivalence classes*

$$[a] = \{b \in \mathbb{Z} : b \equiv a\} = \{a + kn : k \in \mathbb{Z}\} = a + n\mathbb{Z}$$

are referred to as the (*mod n*) **congruence classes** *in $\mathbb{Z}$.*

The quotient space X/R is denoted by $\mathbb{Z}_n$; it consists of the distinct congruence classes in $\mathbb{Z}$. The associated quotient map $\pi = \pi_n : \mathbb{Z} \rightarrow \mathbb{Z}_n$ is given by $\pi : a \rightarrow [a] = a + n\mathbb{Z}$.

Obviously there are exactly n classes, namely

$$[0] = 0 + n\mathbb{Z} = n\mathbb{Z} \qquad [1] = 1 + n\mathbb{Z} \qquad \ldots \qquad [n-1] = (n-1) + n\mathbb{Z}$$

because if we start with some $k \in \mathbb{Z}$ we can add or subtract whole multiples of n to arrive at a (unique) equivalent point k' such that $0 \leq k' < n$. We often describe the classes in $\mathbb{Z}_n$ by choosing particular class representatives. Of course these representatives are not unique, so we could also write $[n-1] = [-1]$, $[n-2] = [-2]$ etc. and sometimes it is useful to do so.

We then showed that the operations $(+)$ and $(\cdot)$ in $\mathbb{Z}$ induce corresponding operations in the quotient space $\mathbb{Z}_n$ of congruence classes, and that the induced operations inherit many properties from $\mathbb{Z}$.

2.4.2 Theorem (Algebraic Structure in the quotient space $\mathbb{Z}_n$). *Fix an integer $n > 1$ Let $\mathbb{Z}_n$ be the quotient space of $(mod\ n)$ congruence classes and let $\pi : \mathbb{Z} \rightarrow \mathbb{Z}_n$ be the quotient map. In $\mathbb{Z}_n$ define operations*

$$(11) \qquad [a] \oplus [b] = [a+b] \qquad \textit{and} \qquad [a] \odot [b] = [ab]$$

for $a, b \in \mathbb{Z}$. These operations are well-defined despite the fact that class representatives are used to define them. They satisfy all the rules set forth in Axioms I*, with*

(a) *The element $[0]$ is the zero element with respect to the $\oplus$ operation*

(b) *The element $[1]$ is the multiplicative identity element with respect to the $\odot$ operation.*

Furthermore the quotient map $\pi : \mathbb{Z} \rightarrow \mathbb{Z}_n$ intertwines all operations, in the sense that

$$(12) \qquad \pi(a+b) = \pi(a) \oplus \pi(b) \qquad \textit{and} \qquad \pi(a \cdot b) = \pi(a) \odot \pi(b)$$

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

The algebraic structure of $\mathbb{Z}_n$ equipped with the $(+)$ operation is pretty simple, but the properties of the multiplication operation $(\,\cdot\,)$ are more subtle. Our intent in this section is to take a closer look at the set of *units* U_n in $\mathbb{Z}_n$. Recall that an element $[a] \in \mathbb{Z}_n$ has a **multiplicative inverse** if there exists some $[k] \in \mathbb{Z}_n$ such that $[k] \cdot [a] = [a] \cdot [k] = [1]$. If it exists this inverse, or "reciprocal," is denoted by $[a]^{-1}$. The invertible elements in $\mathbb{Z}_n$ are the **units** of the system $(\mathbb{Z}_n, +, \cdot)$; we denote them by U_n. The zero element $[0]$ certainly cannot be a unit, and U_n always contains the elements $[1]$ and $-[1] = [-1] = [n-1]$, which might be the *only* units.

In the exercises of Section 1.5 we showed by way of specific examples that the nature of the units U_n varies greatly with the modulus n.

- In $\mathbb{Z}_5$ and $\mathbb{Z}_7$ *every* nonzero element has a multiplicative inverse, so the units in these systems are $\mathrm{U}_5 = \mathbb{Z}_5^\times = \{[x] \in \mathbb{Z}_5 : [x] \neq [0]\}$ and similarly $\mathrm{U}_7 = \mathbb{Z}_7^\times$.
- In $\mathbb{Z}_{12}$ there are only four units $[1], [5], [7], [11]$. Notice that $[11] = [-1] = -[1]$ and $[7] = -[5]$. This system also has zero divisors: $[4] \cdot [3] = [12] = [0]$ even though $[3] \neq [0]$ and $[4] \neq [0]$.

Nevertheless there are important general results concerning units, valid for all moduli. The first asserts that the system of units is closed under the multiplication operation operation $(\,\cdot\,)$ in $\mathbb{Z}_n$

***2.4.3 Exercise.** Verify that the set of units U_n in $\mathbb{Z}_n$ is *closed under multiplication:*

(a) If $[a], [b] \in \mathrm{U}_n$, then $[a][b] = [ab]$ also

(b) If $[a]$ is a unit show that $-[a]$ is a unit and find its inverse. $\square$

***2.4.4 Exercise.** Determine the set of units U_n when:

(a) $n = 4$, (b) $n = 7$, (c) $n = 12$.

Are any of these systems "closed" under the $(+)$ operation? $\square$

2.4.5 Exercise. If an element $[a] \in \mathbb{Z}_n$ has a multiplicative inverse, prove that this inverse element is unique – i.e. if $[u] \cdot [a] = [1]$ and $[u'] \cdot [a] = [1]$, then $[u'] = [u]$. $\square$

The set of units $(\mathrm{U}_n, \cdot\,)$ equipped with multiplication as an operation mapping $\mathrm{U}_n \times \mathrm{U}_n \rightarrow \mathrm{U}_n$ is a new kind of algebraic structure, a **group**. We will spend quite a bit of time discussing groups in the next chapter of these *Notes*, so we only mention the axioms they satisfy.

GROUP AXIOMS. The system $(\mathrm{U}_n, \cdot\,)$ has the following algebraic properties.

(a) ASSOCIATIVITY: $[k]([\ell][m]) = ([k][\ell])[m]$ for all $[k], [\ell], [m]$.

(b) IDENTITY ELEMENT: There exists an element $[1]$ such that $[1][a] = [a][1] = [a]$ for all $[a]$

(c) INVERSES EXIST: For every $[a] \in \mathrm{U}_n$ there exists an element $[b]$ such that $[a][b] = [b][a] = [1]$.

Any system $(G, \cdot\,)$ satisfying these axioms is called a *group*. The particular $G = (\mathrm{U}_n, \cdot\,)$ is a *commutative group* because $[a][b] = [b][a]$.

It turns out that the units in $\mathbb{Z}_n$ can be determined easily without any need for multiplication tables or trial-and-error calculations. The simplest outcome occurs when n is a prime.

2.4.6 Theorem. *If $n > 1$ the groups of units in $\mathbb{Z}_n$ is*

$$\mathrm{U}_n = \{[k] \in \mathbb{Z}_n : 0 < k < n \text{ and } \gcd(k, n) = 1\} \tag{13}$$

PROOF: Assume $0 < k < n$ and $\gcd(k, n) = 1$. Then there must exist integers r, s such that $rk + sn = 1$. Passing to $\mathbb{Z}_n$, we have $[sn] = [0]$ and $[r][k] = [r][k] + [0] = [1]$, so $[k]$ is a unit.

Conversely, if $0 \leq k < n$ and $[k]$ is a unit, then there will be some $[\ell] \in \mathbb{Z}_n$ such that $[1] = [k][\ell] = [k\ell]$. In particular $[k] \neq [0]$ and hence $0 < k < n$. Furthermore, $k\ell \equiv 1 \pmod{n}$ so there exists some $s \in \mathbb{Z}$ such that $k\ell = 1 + sn$. That means $k\ell + (-s)n = 1$, and hence $\gcd(k, n) = 1$ by 2.3.2. □

2.4.7 Corollary. *If $n > 1$ is an integer then all elements $[a] \neq [0]$ in $\mathbb{Z}_n$ have multiplicative inverses $\Leftrightarrow$ the modulus n is a prime.*

PROOF If $p > 1$ is a prime then $\gcd(k, n) = 1$ for all $1 \leq k \leq p - 1$, so by 2.4.6 all nonzero elements in $\mathbb{Z}_p$ have multiplicative inverses. Conversely if $[k]$ has a multiplicative inverse for all $1 \leq k \leq n$, then $\gcd(k, n) = 1$ for all such k. But if n is a non-prime it will have proper divsiors $1 < a, b < n$ such that $ab = n$, which means that $[a]$ and $[b]$ are zero divisors since $[a], [b] \neq [0]$ while $[a] \cdot [b] = [0]$ in $\mathbb{Z}_n$. No zero divisor can be invertible: for if, say, $[a]$ had an inverse $[c]$ that would mean

$$[0] = [c] \cdot [0] = [c] \cdot ([a] \cdot [b]) = ([c] \cdot [a]) \cdot [b] = [1] \cdot [b] = [b] \neq [0]$$

which is impossible. □

So, if $p > 1$ is a prime the set of units U_p in $\mathbb{Z}_p$ is equal to the set of all nonzero elements $\mathbb{Z}_p^\times$, which has cardinality $|\mathrm{U}_p| = p - 1$.

The proof of 2.4.6 also suggests how one might compute the multiplicative inverse of a unit $[k]$ in $\mathbb{Z}_n$: find integers r, s such that $rk + sn = 1$ (whose existence is guaranteed because $gcd = 1$). Then $rk \equiv 1 \pmod{n}$ and $[k]^{-1} = [r]$. The Extended GCD Algorithm is a fast algorithm for computing such a pair r, s.

2.4.8 Exercise. Find the multiplicative inverses $[a]^{-1}$ for

(a) All $[a] \neq [0]$ in $\mathbb{Z}_7$.

(b) Each of the units $[1], [5], [7], [11]$ in $\mathbb{Z}_{12}$.

(c) Determine the units U_6 in $\mathbb{Z}_6$ and their multiplicative inverses.

Hint: You can save some effort using $[-a]^{-1} = -\big([a]^{-1}\big) = [-1] \cdot [a]^{-1}$. □

2.4.9 Exercise. Identify all elements in $\mathbb{Z}_{18}$ that have multiplicative inverses. Then find the inverse of $[5]$ in $\mathbb{Z}_{18}$ by finding integers r, s such that $5r + 18s = 1$ □

Appendix A: Equivalence of the Induction and Minimum Properties.

In this appendix we prove that the Induction Principle (Ind) in 2.1.11 is equivalent to the Minimum Property (Min) of 2.1.17. The proof that (Ind) $\Rightarrow$ (Min) is fairly straightforward, but the converse requires a subtle choice of the set to which the Induction Principle is to be applied.

A.1 Theorem. *Let* $(R, +, \cdot, >)$ *be a commutative ordered ring with identity, so the conditions in* Axioms I *and* II *are satisfied. Then the following properties are equivalent.*

(a) INDUCTION PRINCIPLE: *If a set* $S \subseteq P_{_R}$ *has the properties* (i) $1 \in S$ *and* (ii) $s \in S \;\Rightarrow\; s+1 \in S$*, then* S *is all of* $P_{_R}$*.*

(b) MINIMUM PROPERTY: *If* S *is a nonempty subset of* $P_{_R}$ *it contains a smallest element* s_0*, such that* $s_0 \in S$ *and* $s_0 \leq s$ *for all* $s \in S$*.*

Before beginning the proof we observe that if (Min) holds in a commutative ordered ring R, then the set $P_{_R}$ of positive elements must itself have a smallest element $x_0 = \min\{P_{_R}\} > 0$. We claim that $x_0 \geq 1$ (and hence $x \geq 1$ for *all* elements in $P_{_R}$ if (Min) holds). Otherwise, by Trichotomy, we would have $0 < x_0 < 1$ and then

$$0 < x_0^2 = x_0 \cdot x_0 < 1 \cdot x_0 = x_0$$

which would mean $x_0^2 > 0$ is smaller than the smallest element in $P_{_R}$. That is impossible.

Then if $x > 0$ and $x \neq 1$ it follows that $x > 1$, which implies that $x \geq 2$ because $x > 1 \Rightarrow x - 1 > 0 \Rightarrow x - 1 \geq 1 \Rightarrow x \geq 1 + 1 = 2$. We now take up the main proof.

PROOF (b) $\Rightarrow$ (a): We argue by contradiction: assuming that (Min) holds but (Ind) fails to be true we will produce a contradiction. If (Ind) fails to be true then there exists a set $S \subseteq P_{_R}$ with properties (i) and (ii), such that $S \neq P_{_R}$. The complementary set $S' = P_{_R} \sim S$ is then a nonempty subset of $P_{_R}$. By (Min) there is a smallest element $s_0 = \min\{S'\}$, which by the preceding remarks must be ≥ 1.

We cannot have $s_0 = 1$ because $1 \in S$ while $s_0 \in S'$ cannot be in S. Thus we must have $s_0 \geq 2$ (as noted above) and $s_0 - 1 \geq 1$. Since s_0 is the smallest element in S', $s_0 - 1$ lies in $P_{_R}$ but not in S'. But by definition of S and S' we have $P_{_R} = S \cup S'$ (disjoint union), so if $s_0 - 1$ is not in S' it must lie in S. Then by definition of S (property (ii)), the successor $s_0 = (s_0 - 1) + 1$ is also in S. That's impossible since this successor s_0 lies in $P_{_R} \sim S$. We conclude that $\neg(\text{Ind}) \wedge (\text{Min})$ cannot be true, and hence its negation

$$\neg(\neg(\text{Ind}) \wedge (\text{Min})) \equiv (\text{Ind}) \vee \neg(\text{Min}) \equiv (\,(\text{Min} \Rightarrow (\text{Ind})\,)$$

must hold. That is what we wanted to prove.

PROOF: (a) $\Rightarrow$ (b): Since (Ind) is assumed to hold, R is the usual system of integers $\mathbb{Z}$ and our previous remarks 2.1.12 - 2.1.15 are available. Thus $n > 0$ automatically implies that $n \geq 1$ and may define the *interval* $[1, n] = \{y \in P_{_R} : 1 \leq y \leq n\}$. Consider the following set $A \subseteq P_{_R}$, to which we will apply the Induction Property

(14) $$A = \{n \in P_{_R} : \text{every nonempty subset of } [1, n] \text{ has a smallest element}\}$$

Obviously $1 \in A$ because $[1, 1] = \{1\}$ and the only nonempty subset is $S = \{1\}$, whose smallest element is $\min\{S\} = 1$.

Next, A has the inductive property: if $a \in A$ then $a + 1 \in A$. First note that $[1, a+1] = [1, a] \cup \{a+1\}$ (disjoint union) by 2.1.14. Therefore if S is a nonempty subset of $[1, a+1]$ there are two possibilities:

- $S \cap [1, a] = \emptyset$, in which event $S = \{a + 1\}$ and $\min\{S\} = a + 1$.
- $S' = S \cap [1, a]$ *is nonempty.* Since we are assuming $a \in A$ and $S' \subseteq [1, a]$, the smallest element $s'_0 = \min\{S'\}$ must exist by the Induction Hypothesis. But $s'_0 \leq a < a + 1$, so even if $a + 1 \in S$ we still have $\min\{S\} = \min\{S'\} = s'_0$.

We conclude that any nonempty subset $S \subseteq [1, a{+}1]$ has a smallest element, and hence $a{+}1 \in A$. Since the Induction Principle is valid in R we conclude that A is equal to all of $P_{_R} = \mathbb{N}$, so (14) holds for every positive integer n.

Now consider an *arbitrary* nonempty subset S in $P_{_R}$. Pick some element $x_0 \in S$. Then $S' = S \cap [1, x_0]$ is nonempty; by (14) $x_0 \in A$, which implies that $s_0 = \min\{S \cap [1, x_0]\}$ exists. But all points $y \in S$ lying outside of the interval $[1, x_0]$ must be larger than x_0, so $s_0 \leq x_0 < y$. Thus $s_0 \leq y$ for *all* $y \in S$ and $s_0 = \min\{S\}$. That concludes the proof of the theorem. $\square$

Appendix B: The Schroeder-Bernstein Theorem.

Here we provide a proof of the Schroeder-Bernstein Theorem. The discussion is keyed to the situation shown in Figure 2.1 below.

Figure 2.1. The steps involved in proving the Schroeder-Bernstein Theorem.

B.1. Theorem (Schroeder-Bernstein). *Let A, B be any sets, finite or not, and suppose that*

(i) *$A \approx B'$ for some subset $B' \subseteq B$, so there is some one-to-one map $f : A \rightarrow B$ whose range is B'.*

(ii) *$B \approx A'$ for some subset $A' \subseteq A$, so there is some one-to-one map $g : B \rightarrow A$ whose range is A'.*

Then there exists a bijection $A \approx B$.

Proof: As shown in Figure 2.1, we consider the recursively defined decreasing families of image sets

$$A' = A'_1 = g(B) \supseteq A'_2 = g(B'_1) \supseteq A'_3 = g(B'_2) \supseteq \ldots \supseteq A'_\infty = \bigcap_{n=1}^{\infty} A'_n \quad \text{(the } \textit{residual set} \text{ in } A\text{)}$$

$$B' = B'_1 = f(A) \supseteq B'_2 = f(A'_1) \supseteq B'_3 = f(A'_2) \supseteq \ldots \supseteq B'_\infty = \bigcap_{n=1}^{\infty} B'_n \quad \text{(the } \textit{residual set} \text{ in } B\text{)}$$

The disjoint sets $A_1, A_2, \ldots$ and $B_1, B_2, \ldots$ are defined as set-theoretic differences, taking $A_1 = A \sim A'_1, A_2 = A'_1 \sim A'_2, \ldots$ and so on. There are obvious bijections between *pairs* $A_{2n-1} \cup A_{2n}$ and $B_{2n-1} \cup B_{2n}$; they are obtained by taking

$$\begin{cases} \psi : A_{2n-1} \rightarrow B_{2n} & \text{with} \quad \psi = f \quad \text{on} \quad A_{2n-1} \\ \psi : A_{2n} \rightarrow B_{2n-1} & \text{with} \quad \psi = g^{-1} \quad \text{on} \quad B_{2n-1} \end{cases}$$

This is illustrated in Figure 2.1. We get a bijection between the subsets $A \sim A'_\infty$ and $B \sim B'_\infty$.

As for the "residual sets" (heavily shaded in Figure 2.1), we now show that $f : A'_\infty \rightarrow B'_\infty$ is already a bijection, so we may define $\psi = f$ there to get the desired bijection $\psi : A \rightarrow B$. Our introductory remarks on set theory stressed the fact that an arbitrary map $\phi : X \rightarrow Y$ might fail to preserve intersections of sets (with $\phi(A_1 \cap A_2) \neq \phi(A_1) \cap \phi(A_2)$), although it always respects union of sets (so $\phi\big(\bigcup_{n=1}^{\infty} A_n\big) = \bigcup_{n=1}^{\infty} \phi(A_n)$). One exception occurs when ϕ is one to one; then both intersections and unions are preserved. This applies in the present situation because f and g are one-to-one maps.

It suffices to show $f(A'_\infty) = B'_\infty$. By the preceding remark

$$f\big(\bigcap_{n=1}^{\infty} A'_n\big) = \big(\bigcap_{n=1}^{\infty} f(A'_n)\big) = \bigcap_{n=1}^{\infty} B'_{n+1}$$

Since the family of sets $\{B'_m\}$ is *decreasing* with n, the last intersection is equal to $\bigcap_{n=1}^{\infty} B'_n = B'_\infty$, as required. That completes the proof. □

Note that we might have to perform a countably infinite set of slicing-and-dicing operations on the original maps f and g to construct the final map ψ.