Contents

1 The Integers

The most fundamental example of a ring is the set of integers. Our goal in this chapter is to define the integers axiomatically and to develop some basic properties of primes, divisibility, and modular arithmetic: these will serve as a prototype for much of our analysis of general rings in subsequent chapters.

1.1 The Integers, Axiomatically

- We are all quite familiar with the integers $\mathbb{Z}$, consisting of the natural numbers $\mathbb{N}$ (1, 2, 3, 4, ...), along with their negatives (−1, −2, −3, −4, ...) and zero (0).
    - But it is not quite so simple to prove things about the integers without a solid set of properties to work from!

1.1.1 Axioms for the Integers

- In order to put everything on rigorous ground, we define the integers to be a set $\mathbb{Z}$ along with two (closed) binary[1] operations + and ·, obeying the following properties:

[A1] The operation + is associative: $a + (b + c) = (a + b) + c$ for any integers a, b, c.

[A2] The operation + is commutative: $a + b = b + a$ for any integers a, b.

[A3] There is an additive identity 0 satisfying $a + 0 = a$ for all integers a.

[A4] Every integer a has an additive inverse $-a$ satisfying $a + (-a) = 0$.

[M1] The operation · is associative: $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for any integers a, b, c.

[1] The definition of a binary operation means that for any two integers a and b, the symbols $a + b$ and $a \cdot b$ are always defined and are integers. Some authors list these properties explicitly as part of their list of axioms.

[M2] The operation $\cdot$ is commutative: $a \cdot b = b \cdot a$ for any integers a, b.

[M3] The operation $\cdot$ distributes over $+$: $a \cdot (b + c) = a \cdot b + a \cdot c$ for any integers a, b, c.

[M4] There is a multiplicative identity $1 \neq 0$ satisfying $1 \cdot a = a$ for all integers a.

Furthermore, there is a subset of $\mathbb{Z}$, called $\mathbb{N}$, such that

[N1] For every $a \in \mathbb{Z}$, precisely one of the following holds: $a \in \mathbb{N}$, $a = 0$, or $(-a) \in \mathbb{N}$.

[N2] $\mathbb{N}$ is closed under $+$ and $\cdot$.

[N3] Every nonempty subset S of $\mathbb{N}$ contains a smallest element: that is, an element $x \in S$ such that if $y \in S$, then either $y = x$ or $y - x \in \mathbb{N}$.

- Notation: We often omit the dot symbol for multiplication and write ab in place of $a \cdot b$.
- Definition: The axiom (N3) is called the well-ordering principle. It is the axiom that differentiates the integers from other number systems such as the rational numbers or the real numbers (both of which obey all of the other axioms).

1.1.2 Basic Arithmetic in $\mathbb{Z}$

- We would like to use standard notation for integer arithmetic whenever possible. Our immediate goal, therefore, is to establish a number of simple properties of integer arithmetic.
- Definition: We can define the binary operation of subtraction by setting $a - b = a + (-b)$.
- Definition: Using the definition of $\mathbb{N}$, we can define a relation "$<$" by saying $a < b$ if and only if $b - a \in \mathbb{N}$. (We define $b > a$ to be the same thing.)
  - The axioms [N1] and [N2] ensure that this symbol behaves in the way we would expect an inequality symbol to behave: for any a and b, exactly one of $a < b$, $a = b$, or $b < a$ holds, $a < b$ and $b < c$ imply $a < c$, and $a < b$ with $0 < c$ implies $ac < bc$.
  - We can also define the "non-strict" inequality symbol: if $a < b$ or $a = b$, we say that $a \leq b$.
- Using the axioms for $\mathbb{Z}$, we can establish a multitude of properties of basic arithmetic. Doing this is not especially difficult: it typically requires applying a few of the axioms in creative ways, often in tandem with some case analysis.
- Proposition (Basic Arithmetic): In the integers $\mathbb{Z}$, the following are true:

1. The additive identity 0 is unique, as is the multiplicative identity 1.
   - Proof: Suppose that 0_a and 0_b were both additive identities. Then by [A2] and the hypotheses, $0_a = 0_a + 0_b = 0_b + 0_a = 0_b$. An analogous argument with [M2] shows that the multiplicative identity is unique.
2. Addition has a cancellation law: if $a + b = a + c$, then $b = c$.
   - Proof: By [A1]-[A4], $b = 0 + b = [(-a) + a] + b = (-a) + [a + b] = (-a) + [a + c] = [(-a) + a] + c = 0 + c = c$.
3. Additive inverses are unique.
   - Proof: Suppose that b and c were both additive inverses of a. Then $a + b = 0 = a + c$, so by property (2), $b = c$.
4. For any integer a, $0 \cdot a = 0$.
   - Proof: By [A3], [M3] and [M4], we have $a + 0 = a = 1 \cdot a = (1 + 0) \cdot a = 1 \cdot a + 0 \cdot a = a + 0 \cdot a$. Then by property (2), we conclude $0 \cdot a = 0$.
5. For any integer a, $-(-a) = a$.
   - Proof: By definition, $-(-a)$ has the property that $(-a) + [-(-a)] = 0$. But by [A2] applied to [A4], we also know $(-a) + a = 0$, so by property (3), we conclude $-(-a) = a$.

6. For any integer a, $(-1) \cdot a = -a$.
   - Proof: By [A4], [M3], and the previous property, we have $0 = 0 \cdot a = [1 + (-1)] \cdot a = 1 \cdot a + (-1) \cdot a = a + (-1) \cdot a$. Therefore, $(-1) \cdot a$ is an additive inverse of a, so by property (3), we see $(-1) \cdot a = -a$.
7. For any integers a and b, $-(a + b) = (-a) + (-b)$.
   - Proof: By property (6) and [M3], $-(a + b) = (-1) \cdot (a + b) = (-1) \cdot a + (-1) \cdot b = (-a) + (-b)$.
8. For any integers a and b, $(-a) \cdot b = -(a \cdot b) = a \cdot (-b)$, and $(-a) \cdot (-b) = a \cdot b$.
   - Proof: Observe that $a \cdot b + (-a) \cdot b = [a + (-a)] \cdot b = 0 \cdot b = 0$ by [A4], [M3], and property (4). Thus, $(-a) \cdot b$ is an additive inverse of $a \cdot b$, so by property (3), it is equal to $-(a \cdot b)$. A similar argument shows that $a \cdot (-b) = -(a \cdot b)$. For the last statement, observe that $(-a) \cdot (-b) = -[a \cdot (-b)] = -(-[a \cdot b]) = a \cdot b$ by the first two statements and property (5).
9. The multiplicative identity $1 \in \mathbb{N}$.
   - Proof: By [N1], either $1 \in \mathbb{N}$ (in which case we are done), $1 = 0$ (impossible by [M4]), or $-1 \in \mathbb{N}$. If $-1 \in \mathbb{N}$, then since $\mathbb{N}$ is closed under multiplication by [N2], and since $(-1) \cdot (-1) = 1$ by property (8), we would again conclude that $1 \in \mathbb{N}$. Thus, $1 \in \mathbb{N}$.
10. If $ab = 0$, then $a = 0$ or $b = 0$. If $ab > 0$ then either $a > 0$ and $b > 0$, or $a < 0$ and $b < 0$.
    - Proof: Analyze each of the nine possible cases for whether a or b is positive, negative, or zero, using property (8) and [N1] to determine whether $ab < 0$, $ab = 0$, or $ab > 0$ in each case.
11. If $a < b$ and $b < c$ then $a < c$.
    - Proof: Since $a > b$ we know $b - a$ is in $\mathbb{N}$, and since $b < c$ we know $c - b$ is in $\mathbb{N}$. Then their sum $(c - b) + (b - a) = c - a$ is also in $\mathbb{N}$ by [N2].
12. If $b < c$ then $a + b < a + c$, and also if $a > 0$ then $ab < ac$.
    - Proof: Note that $b < c$ is the same as saying $(c - b) \in \mathbb{N}$, and since $(a + c) - (a + b) = c - b$ by [A1]-[A4] and property (7), we conclude $a + b < a + c$. Also, if $a > 0$ then $a \cdot (c - b) = a \cdot c - a \cdot b$ is also in $\mathbb{N}$ by [N2], so $ab < ac$.

- By judicious application of the arithmetic properties above, we can justify most basic algebraic statements and notation.
  - From this point onward, we will revert to using standard algebraic notation and properties without justifying each individual step, since it is incredibly tedious to write proofs relying solely on axiomatic calculations like the ones given above.

- A rather obvious yet bizarrely important property of the integers is the following result, whose proof we include separately:

- Proposition: There are no integers between 0 and 1.
  - Proof: Let S be the collection of all integers between 0 and 1. If S is empty, we are done, so now assume S is nonempty.
  - By the well-ordering principle, S has a minimal element r.
  - Now observe that since $0 < r < 1$, it is true that $0 < r^2 < r < 1$. But this is a contradiction, because r^2 is then a positive integer less than r, but r was assumed to be minimal.

- Using this property we can establish a few more results, such as the following:

- Proposition: If $ab = 1$, then $a = b = 1$ or $a = b = -1$.
  - Proof: From the basic arithmetic properties, we know that if $a = b = 1$ or $a = b = -1$ then $ab = 1$.
  - Also, since $ab > 0$ we must have $a > 0$ and $b > 0$ or $a < 0$ and $b < 0$.
  - If $a > 0$ and $b > 0$, then $a \geq 1$ and $b \geq 1$, since there are no integers between 0 and 1.
  - Then if $a > 1$ we would have $ab > b \geq 1$, so $ab > 1$. Likewise, if $b > 1$ then $ab > a \geq 1$, so again $ab > 1$.
  - In a similar way, if $a < 0$ and $b < 0$ then $a \leq -1$ and $b \leq -1$, and if $a < -1$ or $b < -1$ then $ab > 1$.

1.1.3 Induction

- As an application of our treatment of the integers, we can establish the validity of "proof by induction".
- The principle of mathematical induction is as follows: suppose we have a sequence of statements $P(1)$, $P(2)$, $P(3)$, and so forth. If $P(1)$ is true, and $P(n)$ implies $P(n+1)$ for every $n \geq 1$, then $P(k)$ is true for every positive integer k.
  - A useful analogy for understanding this inductive principle is of climbing a ladder: if we can get on the first rung of the ladder, and we can always climb from one rung to the next, then we can eventually climb to any rung of the ladder (no matter how high).
  - We often refer to the step of showing that $P(1)$ is true as the base case, and the step of showing that $P(n)$ implies $P(n+1)$ for every $n \geq 1$ as the inductive step.
- Proposition ("Proof by Induction"): If S is a set of positive integers such that $1 \in S$, and $n \in S$ implies $(n+1) \in S$, then $S = \mathbb{N}$.
  - Proof: Let T be the set of elements of $\mathbb{N}$ not in S. If T is empty, we are done, so assume T is nonempty.
  - By the well-ordering principle, T has a minimal element r.
  - Since r is positive, there are three possibilities: $0 < r < 1$, $r = 1$, or $1 < r$.
  - Since there are no positive integers between 0 and 1, and $1 \in S$, the only remaining possibility is that $1 < r$. But then $0 < r - 1$, so $r - 1$ is a positive integer. Since $r - 1 < r$ and r is minimal, we see that $r - 1 \in S$. But the hypotheses on S then imply $r \in S$, which is a contradiction since we assumed $r \in T$.
  - Hence T must be empty, so $S = \mathbb{N}$.
- If we let S be the set of positive integers n such that a statement $P(n)$ holds, then by the result above, if $1 \in S$ and $n \in S$ implies $(n+1) \in S$, then $S = \mathbb{N}$: in other words, if $P(1)$ is true and $P(n)$ implies $P(n+1)$ for every $n \geq 1$, then $P(k)$ is true for every positive integer k.
- Example: Prove that $1 + 3 + 5 + \cdots + (2n-1) = n^2$ for every positive integer n.
  - We prove this by induction on n.
  - For the base case $n = 1$, we must show that $1 = 1$ which is clearly true.
  - For the inductive step, we are given that $1 + 3 + 5 + \cdots + (2n-1) = n^2$ and must show that $1 + 3 + 5 + \cdots + (2n-1) + (2n+1) = (n+1)^2$.
  - By the inductive hypothesis, we can write

$$\begin{aligned} 1 + 3 + 5 + \cdots + (2n-1) + (2n+1) &= [1 + 3 + 5 + \cdots + (2n-1)] + (2n+1) \\ &= n^2 + 2n + 1 = (n+1)^2 \end{aligned}$$

    and therefore we see $1 + 3 + 5 + \cdots + (2n-1) + (2n+1) = (n+1)^2$, as required.
  - By induction, $1 + 3 + 5 + \cdots + (2n-1) = n^2$ for every positive integer n.
- There are various modifications to this "basic" form of induction. The procedure for any induction problem is essentially the same, however: we establish a base case, and prove an inductive step. As long as we establish the appropriate base case and inductive step, the inductive principle will still work.
- Example: Show that $2^n > n^2$ for all integers $n \geq 5$.
  - We prove this by induction on n.
  - We start with the base case $n = 5$: here, we must show that $2^5 > 5^2$, or $32 > 25$, which is true.
  - For the inductive step, we are given that $2^n > n^2$ and $n \geq 5$, and must show that $2^{n+1} > (n+1)^2$.
  - By the inductive hypothesis, we can write $2^{n+1} = 2 \cdot 2^n > 2n^2$.
  - Furthermore, since $n \geq 5$, we have $2n^2 = n^2 + n^2 \geq n^2 + 5n \geq n^2 + 2n + 1 = (n+1)^2$.
  - Putting the inequalities together, we see that $2^{n+1} > 2n^2 \geq (n+1)^2$, so $2^{n+1} > (n+1)^2$ as required.

- Therefore, by induction, $2^n > n^2$ for all integers $n \geq 5$.

- Another flavor of induction is called "complete induction" or "strong induction": rather than assuming the immediately previous case, we assume *all* of the previous cases: the inductive step is now that $P(1)$, $P(2)$, ..., $P(n)$ collectively imply $P(n+1)$.
  - Strong induction makes it easier to construct induction arguments when more than one previous case is needed for proving the inductive step, or when the needed case is not the immediately previous one.
  - This form of induction also follows from our proposition above: simply take S to be the set of positive integers n such that all the statements $P(1)$, $P(2)$, ..., $P(n)$ are true.
  - In fact, (regular) induction and strong induction are equivalent to one another: any proof using one can be rephrased using the other. Thus, it is always possible to start an induction argument with the strong induction hypothesis, even if it is not fully needed during the proof.

- Example: Show that every positive integer n can be written in the form $n = 2^k b$ where $k \geq 0$ and b is odd.
  - We prove this by strong induction on n.
  - For the base case $n = 1$, we can take $k = 0$ and $b = 1$.
  - For the inductive step, now suppose that $n \geq 2$ and that every positive integer less than n has the required property.
  - If n is odd, then we can take $k = 0$ and $b = n$.
  - If n is even, then since $n/2$ is a positive integer less than n, we can write $n/2 = 2^k b$ for some $k \geq 0$ and odd b.
  - Then $n = 2^{k+1} b$ can also be written in the desired form, as required.

1.2 Divisibility and the Euclidean Algorithm

- We have constructed three of the operations of standard arithmetic (namely $+$, $-$, and $\cdot$): now we now turn our attention to division.
  - However, unlike the first three operations, it is not always possible to divide one integer by another and obtain an integer as a quotient.
  - We will therefore start by discussing divisibility.

1.2.1 Divisibility

- Definition: If $a \neq 0$, we say that a divides b (equivalently, b is divisible by a), written $a|b$, if there is an integer k with $b = ka$.
  - Examples: $2|4$, $(-7)|7$, and $6|0$.

- There are a number of basic properties of divisibility that follow immediately from the definition and properties of arithmetic:
  - If $a|b$, then $a|bc$ for any c.
  - If $a|b$ and $b|c$, then $a|c$.
  - If $a|b$ and $a|c$, then $a|(xb + yc)$ for any x and y.
  - If $a|b$ and $b|a$, then $a = b$ or $a = -b$.
  - If $a|b$, and $a, b > 0$, then $a \leq b$.
  - For any $m \neq 0$, $a|b$ is equivalent to $(ma)|(mb)$.

- If $0 < b < a$ and b does not divide a, we can still attempt to divide a by b to obtain a quotient and remainder: this is a less-explicit version of the long-division algorithm familiar from elementary school. Formally:

- Theorem (Division Algorithm): If a and b are positive integers, then there exist unique integers q and r such that $a = qb + r$ with $0 \leq r < b$. Furthermore, $r = 0$ if and only if $b|a$.
  - Proof: The last statement follows immediately from the first part.
  - To show existence, let T be the intersection of the set $S = \{a + kb,\, k \in \mathbb{Z}\}$ with the positive integers. Observe that since $a \in S$, T is nonempty.
  - Let r be the minimal element of T: then $0 \leq r$, and since $r - b$ is not in T by minimality, we also have $r < b$, so $0 \leq r < b$.
  - Furthermore, since r is in the set S, we must have $r = a - qb$ for some integer q.
  - For uniqueness, suppose $qb + r = a = q'b + r'$ with $0 \leq r, r' < b$.
  - Then $-b < r - r' < b$, but we can write $r - r' = b(q' - q)$, so dividing through by b yields $-1 < q' - q < 1$.
  - But since $q' - q$ is an integer and there are no integers between 0 and 1 (or -1 and 0), it must be the case that $q' = q$ and $r' = r$.
  - Example: If $a = 25$ and $b = 4$, then the set $S = \{-7, -3, 1, 5, \ldots, 21, 25, 29, 33, \ldots\}$, and $T = \{1, 5, 9, \ldots\}$. The minimal element of T is $r = 1$, and then we obtain $q = \dfrac{a - r}{b} = 6$. And indeed, we have $25 = 6 \cdot 4 + 1$.
  - In practice, of course, we would not actually construct the sets S and T to determine q and r: we would just numerically compute $25/4$ and round down to the nearest integer to find q.
- Definition: If $d|a$ and $d|b$, then d is a common divisor of a and b. If a and b are not both zero, then there are only a finite number of common divisors: the largest one is called the greatest common divisor, or gcd, and denoted by $\gcd(a, b)$.
  - Warning: Many authors use the notation (a, b) to denote the gcd of a and b: this stems from the notation used for ideals in ring theory. The author of these notes generally dislikes using this notation and will write gcd explicitly, since otherwise it is easy to confuse the gcd with an ordered pair (a, b).
  - Example: The positive divisors of 30 are 1, 2, 3, 5, 6, 10, 15, 30. The positive divisors of 42 are 1, 2, 3, 6, 7, 14, 21, 42. The common (positive) divisors are 1, 2, 3, and 6, and the gcd is therefore 6.
- Theorem (GCD and Linear Combinations): If $d = \gcd(a, b)$, then there exist integers x and y with $d = ax + by$: in fact, the gcd is the smallest positive number of the form $ax + by$.
  - This theorem says that the greatest common divisor of two integers is an integral linear combination of those integers.
  - Proof: Without loss of generality assume $a \neq 0$, and let $S = \{as + bt : s, t \in \mathbb{Z}\} \cap \mathbb{N}$.
  - Clearly S is nonempty since one of a and $-a$ is in S, so now let $l = ax + by$ be the minimal element of S.
  - We claim that $l|b$.
    - Apply the division algorithm to write $b = ql + r$ for some $0 \leq r < l$.
    - Observe that $r = b - ql = b - q(ax + by) = a(-qx) + b(1 - qy)$ is a linear combination of a and b. It is not negative, but it also cannot be positive because otherwise it would necessarily be less than l, and l is minimal.
    - Hence $r = 0$, so $l|b$.
  - By a symmetric argument, $l|a$, and so l is a common divisor of a and b. Therefore, $l \leq d$.
  - But now since $d|a$ and $d|b$ we can wrote $a = dk_a$ and $b = dk_b$ for some integers k_a and k_b, and then $l = ax + by = dk_a x + dk_b y = d(k_a x + k_b y)$.
  - Therefore $d|l$, so in particular $d \leq l$ since both are positive. Since $l \leq d$ as well, we conclude $l = d$.
- Corollary: If $l|a$ and $l|b$, then l divides $\gcd(a, b)$.
  - Proof: Since $l|a$ and $l|b$, l divides any linear combination of a and b: in particular, it divides the gcd.
- As an example: we saw above that the gcd of 30 and 42 is 6, and indeed we can see that $3 \cdot 30 - 2 \cdot 42 = 6$. The other common divisors are 1, 2, and 3, and indeed they all divide 6.

- As another example: because $6 \cdot 24 - 11 \cdot 13 = 1$, we see that 24 and 13 have greatest common divisor 1, since their gcd must divide any linear combination. Having a gcd of 1 occurs often enough that we give this situation a name:
- Definition: If $\gcd(a, b) = 1$, we say a and b are relatively prime.
- Using these results we can quickly prove a number of useful facts about greatest common divisors:
  - If $m > 0$, then $m \cdot \gcd(a, b) = \gcd(ma, mb)$: we can write
$$\gcd(ma, mb) = \min_{x,y \in \mathbb{Z}} [\{max + mby\} \cap \mathbb{N}] = m \cdot \min_{x,y \in \mathbb{Z}} [\{ax + by\} \cap \mathbb{N}] = m \cdot \gcd(a, b).$$
  - If $d > 0$ divides both a and b, then $\gcd(a/d, b/d) = \gcd(a, b)/d$: simply apply the above result to a/d and b/d with $m = d$, and rearrange.
  - If a and b are both relatively prime to m, then so is ab: there exist x_1, y_1, x_2, y_2 with $ax_1 + my_1 = 1$ and $bx_2 + my_2 = 1$. Multiplying yields $ab(x_1x_2) + m(y_1bx_2 + y_2ax_1 + my_1y_2) = 1$, meaning that ab and m are relatively prime.
  - For any integer x, $\gcd(a, b) = \gcd(a, b + ax)$: the set of linear combinations of a and b is the same as the set of integral linear combinations of a and $b + ax$.
  - If $c|ab$ and b, c are relatively prime, then $c|a$: by the first property listed, $\gcd(ab, ac) = a \cdot \gcd(b, c) = a$. Since $c|ab$ and $c|ac$, we conclude $c|a$.

1.2.2 The Euclidean Algorithm

- One question raised by the previous theorems is: how can we actually compute the gcd, except by actually writing down lists of common divisors? And how can we compute the gcd as a linear combination of the original integers? Both questions have a nice answer:
- Theorem (Euclidean Algorithm): Given integers $0 < b < a$, repeatedly apply the division algorithm as follows, until a remainder of zero is obtained:
$$\begin{aligned} a &= q_1 b + r_1 \\ b &= q_2 r_1 + r_2 \\ r_1 &= q_3 r_2 + r_3 \\ &\vdots \\ r_{k-1} &= q_k r_k + r_{k+1} \\ r_k &= q_{k+1} r_{k+1}. \end{aligned}$$
  Then $\gcd(a, b)$ is equal to the last nonzero remainder, r_{k+1}. Furthermore, by successively solving for the remainders and plugging in the previous equations, r_{k+1} can be explicitly written as a linear combination of a and b.
  - Proof: First observe that the algorithm will eventually terminate, because $b > r_1 > r_2 > \cdots \geq 0$, and the well-ordering principle dictates that there cannot exist an infinite decreasing sequence of nonnegative integers.
  - We now claim that $\gcd(a, b) = \gcd(b, r_1)$: this follows because $\gcd(b, r_1) = \gcd(b, a - q_1 b) = \gcd(b, a)$ from the gcd properties we proved earlier.
  - Now we can repeatedly apply this fact to see that $\gcd(a, b) = \gcd(b, r_1) = \gcd(r_1, r_2) = \cdots = \gcd(r_k, r_{k+1}) = r_{k+1}$ since r_{k+1} divides r_k.
  - The correctness of the algorithm for computing the gcd as a linear combination follows by an easy induction.
- Example: Find the gcd of 30 and 42 using the Euclidean algorithm, and write the gcd explicitly as a linear combination of 30 and 42.

- First, we use the Euclidean algorithm:

$$\begin{aligned} 42 &= 1 \cdot 30 + 12 \\ 30 &= 2 \cdot 12 + 6 \\ 12 &= 2 \cdot 6 \end{aligned}$$

  and since $\boxed{6}$ is the last nonzero remainder, it is the gcd.
- For the linear combination, we solve for the remainders:

$$\begin{aligned} 12 &= 42 - 1 \cdot 30 \\ 6 &= 30 - 2 \cdot 12 = 30 - 2 \cdot (42 - 1 \cdot 30) = 3 \cdot 30 - 2 \cdot 42 \end{aligned}$$

  so we obtain $\boxed{6 = 3 \cdot 30 - 2 \cdot 42}$.

- In the example above, we could simply have written down all the divisors of each number, and computed the gcd by comparing those lists. However, if the numbers are large, this procedure becomes very inefficient in comparison to the Euclidean algorithm.
- Example: Find the gcd of 1598 and 4879 using the Euclidean algorithm, and write the gcd explicitly as a linear combination of 1598 and 4879.
  - First, we use the Euclidean algorithm:

$$\begin{aligned} 4879 - 3 \cdot 1598 &= 85 \\ 1598 - 18 \cdot 85 &= 68 \\ 85 - 1 \cdot 68 &= 17 \\ 68 - 4 \cdot 17 &= 0 \end{aligned}$$

    and so the gcd is $\boxed{17}$.
  - For the linear combination, we solve for the remainders:

$$\begin{aligned} 85 &= && = 1 \cdot 4879 - 3 \cdot 1598 \\ 68 &= 1598 - 18 \cdot 85 && = -18 \cdot 4879 + 55 \cdot 1598 \\ 17 &= 85 - 1 \cdot 68 && = 19 \cdot 4879 - 58 \cdot 1598 \end{aligned}$$

    so we obtain $\boxed{17 = 19 \cdot 4879 - 58 \cdot 1598}$.

- Definition: If $a|l$ and $b|l$, l is a common multiple of a and b. Among all (nonnegative) common multiples of a and b, the smallest such l is called the least common multiple of a and b.
  - Example: The least common multiple of 30 and 42 is 210.
  - The least common multiple is often mentioned in elementary school in the context of adding fractions (for finding the "least common denominator").
- The least common multiple has fewer nice properties than the gcd. It does obey the relation $m \cdot \text{lcm}(a, b) = \text{lcm}(ma, mb)$:
  - Since ma divides $\text{lcm}(ma, mb)$, we can write $\text{lcm}(ma, mb) = mk$ for some integer k. Then $ma|mk$ and $mb|mk$, so a and b both divide k, whence $k \geq l$, where $l = \text{lcm}(ma, mb)$.
  - On the other hand, certainly ma and mb divide ml, so $ml \geq mk$. We must therefore have $l = k$.
- Proposition: If $a, b > 0$, the gcd and lcm satisfy $\gcd(a, b) \cdot \text{lcm}(a, b) = ab$.
  - Thus, if we want to calculate the lcm of two arbitrary integers, we can just compute the gcd using the Euclidean Algorithm, and then apply the result of this proposition to get the lcm.
  - Proof: First suppose a and b are relatively prime, and let l be a common multiple. Since $a|l$ we can write $l = ak$ for some integer k: then since $b|ak$ and $\gcd(a, b) = 1$, we conclude by properties of divisibility that $b|k$, meaning that $k \geq b$ and thus $l \geq ab$. But clearly ab is a common multiple of a and b, so it is the least common multiple.
  - In the general case, let $d = \gcd(a, b)$. Then $\gcd(a/d, b/d) = 1$, so by the above we see that $\text{lcm}(a/d, b/d) = ab/d^2$. Then $\gcd(a, b) \cdot \text{lcm}(a, b) = d \cdot d\,\text{lcm}(a/d, b/d) = ab$, as desired.

1.3 Primes and Unique Factorization

- Definition: If $p > 1$ is an integer, we say it is prime if there is no d with $1 < d < p$ such that $d|p$: in other words, if p has no positive divisors other than 1 and itself. If $n > 1$ is not prime, meaning that there is some $d|n$ with $1 < d < n$, we say n is composite. (The integer 1 is neither prime nor composite.)
    - The first few primes are 2, 3, 5, 7, 11, 13, 17, 19, and so forth.
- Here is a basic fact about prime numbers that in more advanced contexts is often used as the actual definition of a prime number:
- Proposition: The integer $p > 1$ is prime if and only if $p|ab$ implies that $p|a$ or $p|b$.
    - Proof: First suppose p is prime and that $p|ab$. If $p|a$ we are done, so assume that $p \nmid a$. Consider $\gcd(a, p)$: it divides p, hence is either 1 or p, but it is not p because p does not divide a.
    - Therefore, $\gcd(a, p) = 1$, so a and p are relatively prime. Then since $p|ab$ and a, p are relatively prime, we see that $p|b$, as required.
    - Conversely, suppose that $p > 1$ and $p|ab$ implies $p|a$ or $p|b$. If there were a d with $1 < d < p$ such that $d|p$, then by hypothesis $p = dk$ for some $1 < k < p$. Then $p|dk$, but p cannot divide d or k, since d and k are both less than p. This is a contradiction, so there cannot exist such a d, and so p is prime.
- The prime numbers are often called the "building blocks under multiplication", because every positive integer can be written as the product of prime numbers:
- Proposition (Prime Factorization): Every positive integer $n > 1$ can be written as a product of primes (where a "product" is allowed to have only one term).
    - A representation of n as a product of primes is called a prime factorization of n. (For example, a prime factorization of 6 is $6 = 2 \cdot 3$.)
    - Proof: We use strong induction on n. The result clearly holds if $n = 2$, since 2 is prime.
    - Now suppose $n > 2$. If n is prime, we are done, so assume that n is not prime: then n is composite.
    - By definition, there exists a d with $1 < d < n$ such that $d|n$: then n/d is an integer satisfying $1 < n/d < n$.
    - By the strong induction hypothesis, both d and n/d can be written as a product of primes; multiplying these two products then yields n as a product of primes.
- We can in fact strengthen the previous result:
- Theorem (Fundamental Theorem of Arithmetic): Every integer $n > 1$ can be factored into a product of primes, and this factorization is unique up to reordering of the factors.
    - Proof: Suppose n is minimal and has two different factorizations: $n = p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_l$. If any of the primes p_i and q_i were equal, we could cancel the corresponding terms and obtain a smaller n, so $p_1 \neq q_j$ for any j with $1 \leq j \leq l$.
    - But since p_1 is prime and divides $q_1 q_2 \cdots q_l$, by repeated application of the previous proposition we see that p_1 must divide one of $q_1, q_2, \ldots, q_l$: say, q_i. But the only divisors of q_i are 1 and q_i, and p_1 cannot be either of them. This is a contradiction, and we are done.
- To save space, we group equal primes together when actually writing out the canonical prime factorization: thus, $12 = 2^2 \cdot 3$, $720 = 2^2 \cdot 3^2 \cdot 5$, and so forth.
    - More generally, we can write the prime factorization in the generic form $n = \prod_{i=1}^{j} p_i^{n_i}$, where the p_i are some (finite) set of primes and the n_i are their corresponding exponents.
- Proposition (Divisibility and Factorizations): If $a = \prod_{i=1}^{j} p_i^{a_i}$ and $b = \prod_{i=1}^{j} p_i^{b_i}$, then $a|b$ if and only if $a_i \leq b_i$ for each i. In particular, $\gcd(a, b) = \prod_{i=1}^{j} p_i^{\min(a_i, b_i)}$, and $\text{lcm}(a, b) = \prod_{i=1}^{j} p_i^{\max(a_i, b_i)}$.
    - Proof: We observe that if $b = ak$ and $k = \prod_{i=1}^{j} p_i^{k_i}$, then $a_i + k_i = b_i$. Since all exponents are nonnegative, saying that such an integer k exists is equivalent to saying that $a_i \leq b_i$ for all i.

- The statements about the gcd and lcm then follow immediately, since (for example) the exponent of p_i in the gcd is the largest integer that is $\leq a_i$ and $\leq b_i$, which is a more convoluted way of saying the minimum of a_i and b_i.

- One question we might have is: how many primes are there? The most basic answer to this question is that there are infinitely many primes:

- Theorem (Euclid): There are infinitely many prime numbers.

  - Proof: Suppose there are only finitely many prime numbers $p_1, p_2, \ldots, p_k$, and consider $n = p_1 p_2 \cdots p_k + 1$.
  - Since n is bigger than each p_i, n cannot be prime (since it would necessarily have to be on the list).
  - Therefore n is composite. Consider the prime factorization of n: necessarily at least one prime on the list must appear in it: say p_i.
  - Since p_i also divides $p_1 p_2 \cdots p_k$, we see that p_i therefore divides $n - p_1 p_2 \cdots p_k = 1$. But this is a contradiction. Hence there are infinitely many primes.

- It is much harder to give more precise answers to the question of "how many primes are there?". One result in this direction, whose proof is quite difficult, is as follows:

- Theorem (Prime Number Theorem): Let $\pi(n)$ be the number of primes in the interval $[1, n]$. Then $\lim_{n\to\infty} \dfrac{\pi(n)}{n/\log(n)} = 1$, where log denotes the natural logarithm.

  - This theorem says that the number of primes in the interval $[1, n]$ is approximately $n/\log(n)$ for sufficiently large n.

- There are many unsolved problems about prime numbers that have been studied in the branch of mathematics known as number theory.

1.4 Modular Congruences and $\mathbb{Z}/m\mathbb{Z}$

- The ideas underlying modular arithmetic are familiar to anyone who can tell time. For example, 3 hours after 11 o'clock, it is 2 o'clock, even though $3 + 11$ is 14, not 2. Simply put, we identify times that are 12 hours apart as the same time of day.

- Modular arithmetic is a generalization of this "clock arithmetic".

1.4.1 Modular Congruences

- Definition: If m is a positive integer and m divides $b - a$, we say that a and b are congruent modulo m (or equivalent modulo m), and write "$a \equiv b$ (modulo m)".

  - Notation: As shorthand we usually write "$a \equiv b \pmod{m}$", or even just "$a \equiv b$" when the modulus m is clear from the context.
  - The statement $a \equiv b \pmod{m}$ can be thought of as saying "a and b are equal, up to a multiple of m".
  - Observe that if $m|(b - a)$, then $(-m)|(b - a)$ as well, so we do not lose anything by assuming that the modulus m is positive.
  - Example: $3 \equiv 9 \pmod{6}$, since 6 divides $9 - 3 = 6$.
  - Example: $-2 \equiv 28 \pmod{5}$, since 5 divides $28 - (-2) = 30$.
  - Example: $0 \equiv -666 \pmod{3}$, since 3 divides $-666 - 0 = -666$.
  - If m does not divide $b - a$, we say a and b are not congruent mod m, and write $a \not\equiv b \pmod{m}$.
  - Example: $2 \not\equiv 7 \pmod{3}$, because 3 does not divide $7 - 2 = 5$.

- Modular congruences share a number of properties with equalities:

- Proposition (Modular Congruences): For any positive integer m and any integers a, b, c, d, the following are true:

  1. $a \equiv a \pmod{m}$.
     - Proof: By definition of divisibility, m always divides $a - a = 0$.
  2. $a \equiv b \pmod{m}$ if and only if $b \equiv a \pmod{m}$.
     - Proof: By properties of divisibility, $m|(b-a)$ is equivalent to $m|(a-b)$.
  3. If $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m}$, then $a \equiv c \pmod{m}$.
     - Proof: If $m|(b-a)$ and $m|(c-b)$, then m also divides $(c-b)+(b-a) = c-a$, so that $a \equiv c \pmod{m}$.
  4. If $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, then $a + c \equiv b + d \pmod{m}$.
     - Proof: If $m|(b-a)$ and $m|(d-c)$, then m also divides $(b-a)+(d-c) = (b+d)-(a+c)$, so that $a + c \equiv b + d \pmod{m}$.
  5. If $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, then $ac \equiv bd \pmod{m}$.
     - Proof: If $m|(b-a)$ and $m|(d-c)$, then m also divides $d(b-a)+a(d-c) = bd-ac$, so that $ac \equiv bd \pmod{m}$.
  6. If $a \equiv b \pmod{m}$, then $ac \equiv bc \pmod{mc}$ for any $c > 0$.
     - Proof: If $m|(b-a)$, then by properties of divisibility, $(mc)|(bc-ac)$, and so $ac \equiv bc \pmod{mc}$.
  7. If $a \equiv b \pmod{m}$ then $a^k \equiv b^k \pmod{m}$ for any positive integer k.
     - Proof: Induction on k: the base case $k = 1$ is trivial. For the inductive step, if $a^{k-1} \equiv b^{k-1} \pmod{m}$, then by property (5) we see that $a^k \equiv b^k \pmod{m}$.
  8. If $d|m$, then $a \equiv b \pmod{m}$ implies $a \equiv b \pmod{d}$.
     - Proof: If $d|m$ and $m|(b-a)$, then $d|(b-a)$, and so $a \equiv b \pmod{d}$.

- The first three properties above demonstrate that modular equivalence is an equivalence relation.
  - Remark: A binary relation $\sim$ defined on a nonempty set S is called an equivalence relation if it obeys the following three axioms:

    [E1] For any $a \in S$, $a \sim a$.
    [E2] For any $a, b \in S$, $a \sim b$ implies $b \sim a$.
    [E3] For any $a, b, c \in S$, $a \sim b$ and $b \sim c$ implies $a \sim c$.
  - Example: Equality of elements in any set (e.g., equality of real numbers) is an equivalence relation.

1.4.2 Residue Classes and $\mathbb{Z}/m\mathbb{Z}$

- We would now like to study "arithmetic modulo m". To do this, we need to define the underlying objects of study:
- Definition: If a is an integer, the residue class of a modulo m, denoted $\overline{a}$, is the collection of all integers congruent to a modulo m. Observe that $\overline{a} = \{a + km,\ k \in \mathbb{Z}\}$.
- Example: The residue class of 2 modulo 4 is the set $\{\ldots, -6, -2, 2, 6, 10, 14, \ldots\}$, while the residue class of 2 modulo 5 is the set $\{\ldots, -8, -3, 2, 7, 12, 17, \ldots\}$.
- Here are a few fundamental properties of residue classes:
- Proposition (Properties of Residue Classes): Suppose m is a positive integer. Then

  1. If a and b are integers with respective residue classes $\overline{a}$, $\overline{b}$ modulo m, then $a \equiv b \pmod{m}$ if and only if $\overline{a} = \overline{b}$.

- Proof: If $\overline{a} = \overline{b}$, then by definition b is contained in the residue class $\overline{a}$, meaning that $b = a + km$ for some k. Thus, m divides $b - a$, so $a \equiv b \pmod{m}$.
- Conversely, suppose $a \equiv b \pmod{m}$. If c is any element of the residue class $\overline{a}$, then by definition $c \equiv a \pmod{m}$, and therefore $c \equiv b \pmod{m}$.
- Therefore, c is an element of the residue class $\overline{b}$, but since c was arbitrary, this means that $\overline{a}$ is contained in $\overline{b}$.
- By the same argument with a and b interchanged, we see that $\overline{b}$ is also contained in $\overline{a}$, and thus $\overline{a} = \overline{b}$.

2. Two residue classes modulo m are either disjoint or identical.
   - Proof: Suppose that $\overline{a}$ and $\overline{b}$ are two residue classes modulo m. If they are disjoint, we are done, so suppose there is some c contained in both.
   - Then $c \equiv a \pmod{m}$ and $c \equiv b \pmod{m}$, so $a \equiv b \pmod{b}$. Then by property (1), we conclude $\overline{a} = \overline{b}$.
3. There are exactly m distinct residue classes modulo m, given by $\overline{0}$, $\overline{1}$, ... , $\overline{m-1}$.
   - Proof: By the division algorithm, for any integer a there exists a unique r with $0 \leq r < m$ such that $a = qm + r$ with $q \in \mathbb{Z}$.
   - Then $a \equiv r \pmod{m}$, and so every integer is congruent modulo m to precisely one of the m integers $0, 1, \ldots, m-1$.
   - Equivalently, every integer is contained in exactly one of the residue classes $\overline{0}$, $\overline{1}$, ... , $\overline{m-1}$.
   - By property (2), we conclude that there are exactly m distinct residue classes modulo m: $\overline{0}$, $\overline{1}$, ... , $\overline{m-1}$.

- Definition: The collection of residue classes modulo m is denoted $\mathbb{Z}/m\mathbb{Z}$ (read as "$\mathbb{Z}$ modulo $m\mathbb{Z}$").
  - Notation: Many other authors denote this collection of residue classes modulo m as $\mathbb{Z}_m$. We will avoid this notation and exclusively use $\mathbb{Z}/m\mathbb{Z}$ (or its shorthand $\mathbb{Z}/m$), since $\mathbb{Z}_m$ is used elsewhere in algebra and number theory for a different object.
  - By our properties above, $\mathbb{Z}/m\mathbb{Z}$ contains exactly m elements $\overline{0}$, $\overline{1}$, ... , $\overline{m-1}$.
- We can now write down "addition and multiplication" modulo m using the residue classes of $\mathbb{Z}/m\mathbb{Z}$.
  - The fact that $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$ imply $a + c \equiv b + d \pmod{m}$ and $ac \equiv bd \pmod{m}$ tell us that if we want to compute $a + c$ modulo m, then no matter which element b in the residue class of a and which element d in the residue class of c we take, the sum $b + d$ will lie in the same residue class as $a + c$, and the product bd will lie in the same residue class as ac.
  - Thus, everything makes perfectly good sense if we label the residue classes with the integers 0 through $m - 1$ and simply do the arithmetic with those residue classes.
- Definition: The addition operation in $\mathbb{Z}/m\mathbb{Z}$ is defined as $\overline{a} + \overline{b} = \overline{a + b}$, and the multiplication operation is defined as $\overline{a} \cdot \overline{b} = \overline{ab}$.
  - Here are the addition and multiplication tables for $\mathbb{Z}/5\mathbb{Z}$:

$+$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$
$\overline{0}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$
$\overline{1}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{0}$
$\overline{2}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{0}$	$\overline{1}$
$\overline{3}$	$\overline{3}$	$\overline{4}$	$\overline{0}$	$\overline{1}$	$\overline{2}$
$\overline{4}$	$\overline{4}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$

$\cdot$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$
$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$
$\overline{2}$	$\overline{0}$	$\overline{2}$	$\overline{4}$	$\overline{1}$	$\overline{3}$
$\overline{3}$	$\overline{0}$	$\overline{3}$	$\overline{1}$	$\overline{4}$	$\overline{2}$
$\overline{4}$	$\overline{0}$	$\overline{4}$	$\overline{3}$	$\overline{2}$	$\overline{1}$

  - Note that, for example, the statement $\overline{2} + \overline{4} = \overline{1}$ is now perfectly acceptable (and correctly stated with the equals sign): it says that if we take any element in the residue class $\overline{2}$ (modulo 5) and add it to any element in the residue class $\overline{4}$ (modulo 5), the result will always lie in the residue class $\overline{1}$ (modulo 5).

- Here are the addition and multiplication tables for $\mathbb{Z}/4\mathbb{Z}$:

$+$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{0}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{1}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{0}$
$\overline{2}$	$\overline{2}$	$\overline{3}$	$\overline{0}$	$\overline{1}$
$\overline{3}$	$\overline{3}$	$\overline{0}$	$\overline{1}$	$\overline{2}$

$\cdot$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{2}$	$\overline{0}$	$\overline{2}$	$\overline{0}$	$\overline{2}$
$\overline{3}$	$\overline{0}$	$\overline{3}$	$\overline{2}$	$\overline{1}$

- Arithmetic modulo m is commonly described by ignoring residue classes entirely and only working with the integers 0 through $m-1$, with the result of every computation "reduced modulo m" to obtain a result lying in this range.
  - Thus, for example, to compute $3+10$ modulo 12, we would add to get 13 and then "reduce", yielding 1 modulo 12. Similarly, to find $3 \cdot 10$ modulo 12, we compute $3 \cdot 10 = 30$ and then reduce to obtain a result of 6 modulo 12.
  - However, this is a rather cumbersome and inelegant description. This definition is often used in programming languages, where "a mod m", frequently denoted "$a\%m$", is defined to be a *function* returning the corresponding remainder in the interval $[0, m-1]$.
  - Observe that with this definition, it is not true that $(a+b)\%m = (a\%m) + (b\%m)$, nor is is true that $ab\%m = (a\%m) \cdot (b\%m)$, since the sum and product may each exceed m. Instead, to obtain an actually true statement, one would have to write something like $ab\%m = [(a\%m) \cdot (b\%m)]\%m$.
  - In order to avoid such horrible kinds of statements, the best viewpoint really is to think of the statement $a \equiv b \pmod{m}$ as a congruence that is a "weakened" kind of equality, rather than always reducing each of the terms to its residue in the set $\{0, 1, \ldots, m-1\}$.
  - The other reason we adopt the use of residue classes is that they extend quite well to more general settings where we may not have such an obvious set of "representatives".
- The arithmetic in $\mathbb{Z}/m\mathbb{Z}$ shares many properties with the arithmetic in $\mathbb{Z}$ (which should not be surprising, since $\mathbb{Z}/m\mathbb{Z}$ was constructed using $\mathbb{Z}$):
- Proposition (Basic Arithmetic in $\mathbb{Z}/m\mathbb{Z}$): For any positive integer m the following properties of residue classes in $\mathbb{Z}/m\mathbb{Z}$ hold:
  1. The operation $+$ is associative: $\overline{a} + (\overline{b} + \overline{c}) = (\overline{a} + \overline{b}) + \overline{c}$ for any $\overline{a}$, $\overline{b}$, and $\overline{c}$.
  2. The operation $+$ is commutative: $\overline{a} + \overline{b} = \overline{b} + \overline{a}$ for any $\overline{a}$ and $\overline{b}$.
  3. The residue class $\overline{0}$ is an additive identity: $\overline{a} + \overline{0} = \overline{a}$ for any $\overline{a}$.
  4. Every residue class $\overline{a}$ has an additive inverse $-\overline{a}$ satisfying $\overline{a} + (-\overline{a}) = \overline{0}$.
  5. The operation $\cdot$ is associative: $\overline{a} \cdot (\overline{b} \cdot \overline{c}) = (\overline{a} \cdot \overline{b}) \cdot \overline{c}$ for any $\overline{a}$, $\overline{b}$, and $\overline{c}$.
  6. The operation $\cdot$ is commutative: $\overline{a} \cdot \overline{b} = \overline{b} \cdot \overline{a}$ for any $\overline{a}$ and $\overline{b}$.
  7. The operation $\cdot$ distributes over $+$: $\overline{a} \cdot (\overline{b} + \overline{c}) = \overline{a} \cdot \overline{b} + \overline{a} \cdot \overline{c}$ for any $\overline{a}$, $\overline{b}$, and $\overline{c}$.
  8. The residue class $\overline{1}$ is a multiplicative identity: $\overline{1} \cdot \overline{a} = \overline{a}$ for any $\overline{a}$.
     - Proof: For (1), by definition we have $\overline{a} + (\overline{b} + \overline{c}) = \overline{a} + \overline{b+c} = \overline{a + (b+c)}$ and also $(\overline{a} + \overline{b}) + \overline{c} = \overline{a+b} + \overline{c} = \overline{(a+b)+c}$.
     - But by the associative property [A1] in $\mathbb{Z}$, we know that $a + (b+c) = (a+b) + c$, so the associated residue classes are also equal.
     - The other properties follow in a similar way from the corresponding properties of the integers.

1.4.3 Units and Zero Divisors in $\mathbb{Z}/m\mathbb{Z}$

- We will now abuse notation and ignore the distinction between the integer a and the residue class $\overline{a}$ modulo m: we will now simply refer to "integers modulo m" and drop the residue class notation.

- In this new notation, the multiplication table for $\mathbb{Z}/6\mathbb{Z}$ is as follows:

$\cdot$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

- As we saw above, the arithmetic in $\mathbb{Z}/m\mathbb{Z}$ shares many properties with the arithmetic in $\mathbb{Z}$. However, there are some very important differences.
  - For example, if a, b, c are integers with $ab = ac$ and $a \neq 0$, then we can "cancel" a from both sides to conclude that $b = c$.
  - However, this does not always work in $\mathbb{Z}/m\mathbb{Z}$: for example, $2 \cdot 1 = 2 \cdot 4$ modulo 6, but $1 \neq 4$ modulo 6.
  - The issue here is that 2 and the modulus 6 are not relatively prime: 6 divides $2(4-1)$, but 6 does not divide $4 - 1$.
- We can explain the issue using modular congruences:
- Proposition (Modular Cancellation): If $m > 0$ and $d = \gcd(a, m)$, then $ax \equiv ay \pmod{m}$ is equivalent to $x \equiv y \pmod{m/d}$.
  - Proof: First suppose $ax \equiv ay \pmod{m}$. Then there exists an integer k with $a(y-x) = km$, so $\frac{a}{d}(y-x) = \frac{m}{d}k$. Since $\gcd(a/d, m/d) = \gcd(a, m)/d = 1$, we see that $\frac{m}{d}$ divides $y - x$, meaning that $x \equiv y \pmod{m/d}$.
  - For the other direction, suppose $y - x = \frac{m}{d}k$. Then $a(y-x) = \frac{a}{d}mk$, and since $d|a$, we see that m divides $a(y-x)$, meaning that $ax \equiv ay \pmod{m}$.
- By setting $d = 1$ in this proposition, we obtain an extremely useful result:
- Corollary: If a and m are relatively prime, $ax \equiv ay \pmod{m}$ implies $x \equiv y \pmod{m}$.
  - This corollary tells us when we are allowed to "cancel" a from both sides of a congruence modulo m.
  - From the viewpoint of $\mathbb{Z}/m\mathbb{Z}$, this corollary tells us that if a is relatively prime to m, then the residue class $\overline{a}$ in $\mathbb{Z}/m\mathbb{Z}$ has a "multiplicative inverse".
- Definition: We say a is a unit modulo m if it has a multiplicative inverse: that is, if there is some b such that $ab \equiv 1 \pmod{m}$.
  - The multiplicative inverse b is often written as $a^{-1} \pmod{m}$, or even sometimes as $1/a \pmod{m}$.
  - Example: From the multiplication table modulo 5, we see that 1, 2, 3, and 4 all have multiplicative inverses, which are 1, 3, 2, and 4 respectively, but 0 does not. So 1, 2, 3, and 4 are the units modulo 5.
  - Example: Modulo 6, it is straightforward to check that the only units are 1 and 5 (whose multiplicative inverses are 1 and 5 respectively).
  - Example: Modulo 14, the units are 1, 3 (inverse 5), 5 (inverse 3), 9 (inverse 11), 11 (inverse 9), and 13.
  - Remark: Technically, the definition does not require that there be only one such b. However, there can be only one such b: if $ab_1 \equiv 1 \equiv ab_2 \pmod{m}$, then $b_1 \equiv b_1 ab_2 \equiv b_2 \pmod{m}$.
- From the corollary above, we know that the units modulo m are precisely the residue classes relatively prime to m.
- We can use the Euclidean algorithm to compute the multiplicative inverse of a unit: simply apply the Euclidean algorithm to generate x and y with $xa + ym = 1$: then the inverse of a modulo m is x.
- Example: Show that 7 is a unit modulo 52, and then find its multiplicative inverse.

- We apply the Euclidean algorithm:

$$\begin{aligned} 52 &= 7 \cdot 7 + 3 \\ 7 &= 2 \cdot 3 + 1 \end{aligned}$$

  so the gcd is indeed 1.
- Now we have $3 = 52 - 7 \cdot 7$ and then $1 = 7 - 2 \cdot 3 = 7 - 2 \cdot (52 - 7 \cdot 7) = 15 \cdot 7 - 2 \cdot 52$.
- Then $15 \cdot 7 \equiv 1 \pmod{52}$, so $7^{-1} = \boxed{15 \pmod{52}}$.

- Another important property of arithmetic in $\mathbb{Z}$ (as well as in the real or complex numbers) that we often take for granted is that if x, y are such that $xy = 0$, then $x = 0$ or $y = 0$.
  - This property no longer holds for congruences modulo m for arbitrary m: for example, modulo 6 we have $\overline{2} \cdot \overline{3} = \overline{0}$, but $\overline{2} \neq \overline{0}$ and $\overline{3} \neq \overline{0}$. We give such situations a special name:

- Definition: We say that a residue class $\overline{a}$ modulo m is a zero divisor if $\overline{a} \neq \overline{0}$ and there exists a nonzero $\overline{b}$ such that $\overline{a} \cdot \overline{b} = \overline{0}$. (Note in particular that $\overline{0}$ is *not* a zero divisor!)
  - Example: In $\mathbb{Z}/6\mathbb{Z}$, since $\overline{2} \cdot \overline{3} = \overline{4} \cdot \overline{3} = \overline{0}$, the residue classes represented by 2, 3, and 4 are zero divisors.

- Proposition: In $\mathbb{Z}/m\mathbb{Z}$, a unit can never be a zero divisor.
  - Proof: If a were both a unit and a zero divisor, then there would exist b, x such that $ab = 1$ and $ax = 0$, with $x \neq 0$.
  - But then we would have $x = (ab)x = b(ax) = 0$, contradicting the assumption that $x \neq 0$.

- We can now describe the zero divisors modulo m: they are simply the nonzero residue classes that are not units.

- Proposition: An integer a is a zero divisor modulo m if and only if $1 < \gcd(a, m) < m$.
  - Proof: Let $d = \gcd(a, m)$. We break into cases depending on the value of d.
  - If $d = 1$, then a is a unit, and therefore is not a zero divisor.
  - If $d = m$, then $m|a$ meaning that $\overline{a} = \overline{0}$, and 0 is defined not to be a zero divisor.
  - If $1 < d < m$, then $(m/d) \cdot a = m \cdot (a/d) \equiv 0 \pmod{m}$, and m/d is nonzero. Therefore, a is a zero divisor.

- As a final observation, we remark that when the modulus is a prime p, there are no zero divisors modulo p and every nonzero residue class is a unit. (We will return to study the structure of $\mathbb{Z}/p\mathbb{Z}$ further in a subsequent chapter.)

Well, you're at the end of my handout. Hope it was helpful.