Contents

2 The Integers and Modular Arithmetic

One of the most foundational objects in mathematics is the integers, as they are used the basis and reference point for many other topics in mathematics. Our goal in this chapter is to define the integers axiomatically and to develop some basic properties of divisibility, common divisors, primes and factorizations, and modular arithmetic as a way of illustrating a variety of proof techniques and ideas in a familiar context.

2.1 The Integers, Axiomatically

- We are all at least a little bit familiar with the integers $\mathbb{Z}$, consisting of the positive integers $\mathbb{Z}_+$ (1, 2, 3, 4, ...), along with their negatives (−1, −2, −3, −4, ...) and zero (0).
    - There are two natural binary arithmetic operations defined on the integers, namely addition (+) and multiplication ($\cdot$), along with the unary operation of negation (−).
    - But it is not quite so simple to prove things about the integers without a solid set of properties to work from.

2.1.1 Definition of the Integers

- "Definition": The integers are a set $\mathbb{Z}$ along with two (closed) binary[1] operations $+$ and $\cdot$, obeying the following properties[2]:

  [I1] The operation $+$ is associative: $a + (b + c) = (a + b) + c$ for any integers a, b, c.

  [I2] The operation $+$ is commutative: $a + b = b + a$ for any integers a, b.

  [I3] There is an additive identity 0 satisfying $a + 0 = a$ for all integers a.

  [I4] Every integer a has an additive inverse $-a$ satisfying $(-a) + a = 0$.

  [I5] The operation $\cdot$ is associative: $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for any integers a, b, c.

  [I6] The operation $\cdot$ is commutative: $a \cdot b = b \cdot a$ for any integers a, b.

  [I7] There is a multiplicative identity $1 \neq 0$ satisfying $1 \cdot a = a$ for all integers a.

  [I8] The operation $\cdot$ distributes over $+$: $a \cdot (b + c) = a \cdot b + a \cdot c$ for any integers a, b, c.

  Furthermore, there is a subset of $\mathbb{Z}$, namely the positive integers $\mathbb{Z}_+$, such that

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

  [N2] The set $\mathbb{Z}_+$ is closed under $+$ and $\cdot$: for any $a, b \in \mathbb{Z}_+$, both $a + b$ and $a \cdot b$ are in $\mathbb{Z}_+$.

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

- Remark: The axiom [N3] is called the well-ordering axiom. 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.

2.1.2 Basic Arithmetic

- Using the axioms for $\mathbb{Z}$, we can establish all of the properties of basic arithmetic. Doing this is not especially difficult once the basic idea is identified (namely, invoking the axioms judiciously, along with some case analysis). Here are some examples:

- Proposition (Basic Arithmetic): Inside the integers $\mathbb{Z}$, the following properties hold:

  1. The additive and multiplicative identities are unique.
     - Proof: Suppose we had two additive identities 0_A and 0_B. Then by axioms [I2] and [I3], we may write $0_A = 0_A + 0_B = 0_B + 0_A = 0_B$, and therefore $0_A = 0_B$.
     - In a similar way, if we had two multiplicative identities 1_A and 1_B, then by axioms [I6] and [I7], we may write $1_A = 1_A \cdot 1_B = 1_B \cdot 1_A = 1_B$, and therefore $1_A = 1_B$.
  2. Addition possesses a cancellation law: if $a + b = a + c$, then $b = c$.
     - Proof: By axioms [I1], [I3], and [I4], we have $b = 0 + b = [(-a) + a] + b = (-a) + (a + b) = (-a) + (a + c) = [(-a) + a] + c = 0 + c = c$.
  3. Additive inverses are unique.
     - Proof: Suppose a had two additive inverses b and c. Then we would have $a + b = 0 = a + c$ by [I2] and [I4], and therefore by the cancellation law (2) we would have $b = c$.
  4. For all $a \in \mathbb{Z}$, $0 \cdot a = 0$, $(-1) \cdot a = -a$, and $-(-a) = a$.
     - Proof: For any element a, by [I3] and [I8] we have $0 \cdot a + 0 = 0 \cdot a = (0 + 0) \cdot a = 0 \cdot a + 0 \cdot a$. Then by the cancellation law (2), we obtain $0 = 0 \cdot a$.
     - For the second statement, by the above along with [I3], [I7], and [I8] we have $0 = 0 \cdot a = [1 + (-1)] \cdot a = 1 \cdot a + (-1) \cdot a = a + (-1) \cdot a$. Then by the uniqueness of additive inverses (3), we see $(-1) \cdot a = -a$.

[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.

[2] To be a proper definition, we would also need to establish that there actually is a set with operations obeying these properties, which turns out to be rather tedious. But there are various constructions for $\mathbb{Z}$ using set theory, which we will not detail here.

- For the last statement, observe that by definition, $-(-a)$ is the element which when added to $-a$ yields 0. But since $a+(-a)=0=(-a)+a$ by definition and [I2], by the uniqueness of the additive inverse (3) we conclude $-(-a)=a$.

5. For any a and b, $-(a+b)=(-a)+(-b)$, $(-a)\cdot b=-(a\cdot b)=a\cdot(-b)$, and $(-a)\cdot(-b)=a\cdot b$.
   - Proof: For the first statement, observe that by [I1]-[I4], we have $[a+b]+[(-a)+(-b)]=[a+[b+(-b)]]+(-a)=(a+0)+(-a)=a+(-a)=0$, and so by the uniqueness of the additive inverse (3) we see $-(a+b)=(-a)+(-b)$.
   - For the second statement, by (4) and [I6] we have $(-a)\cdot b=[(-1)\cdot a]\cdot b=(-1)\cdot(a\cdot b)=-(a\cdot b)$. By essentially the same argument the other part of the second statement follows, as does the third statement.
6. For any a and b, $b\cdot a=a+(b-1)\cdot a$. Thus, $2\cdot a=a+a$, $3\cdot a=a+(a+a)$, and so forth.
   - Proof: We have $b\cdot a=[1+(b-1)]\cdot a=1\cdot a+(b-1)\cdot a=a+(b-1)\cdot a$ by [I7] and [I8].
   - The second statement follows from this along with the observation that $1\cdot a=a$.
7. The multiplicative identity $1\in\mathbb{Z}_+$.
   - Proof: By [N1], one of the following things holds: either $1\in\mathbb{Z}_+$ (in which case we are done), or $1=0$ (this is impossible because by [I7], $1\neq 0$), or $-1\in\mathbb{Z}_+$.
   - If $-1\in\mathbb{Z}_+$, then by [N2], we would see that $(-1)\cdot(-1)\in\mathbb{Z}_+$, and by (6) we have $(-1)\cdot(-1)=1$, so we would have $1\in\mathbb{Z}_+$. In all cases $1\in\mathbb{Z}_+$ so we are done.
8. If $ab=0$, then $a=0$ or $b=0$.
   - Proof: If $a,b\in\mathbb{Z}_+$ then $ab\in\mathbb{Z}_+$ and so $ab\neq 0$. If $a,-b\in\mathbb{Z}_+$ or $-a,b\in\mathbb{Z}_+$ then $-(ab)\in\mathbb{Z}_+$ by (5) and (4), and if $-a,-b\in\mathbb{Z}_+$ then $ab\in\mathbb{Z}_+$ also by (5) and (4).
   - Thus, the only case in which $ab=0$ is the case where $a=0$ or $b=0$, as claimed.

- It is quite tedious to write every proof using only properties of the axioms, so from this point forward we will revert to using more standard notation and language.
  - However, it is worthwhile noting that we could (if we wanted to) always reduce every proof down to a series of statements each of which is an application of one of the axioms.
  - From this viewpoint, our intermediate results (our propositions, lemmas, theorems, and so forth) consist of a sequence of applications of the axioms that we can invoke in any situation where the hypotheses apply, and that yield the claimed result.
  - In this way, we can "build up" from the axiomatic foundation, by first proving very basic properties, and then using those results to prove more complicated properties, and so forth, until we have established substantial results.
  - As a matter of course, most mathematicians do not dwell much on foundational questions, and instead take for granted all of the basic properties of numbers and arithmetic that we will examine closely.
  - But, at least in principle, every mathematical proof can be reduced down to a sequence of axiomatic calculations. This idea is actually the foundation of automated theorem provers, which are computer programs that can construct and verify mathematical proofs down to the axiomatic level.
  - We cluster these statements together to make them more readable and (vastly!) more understandable to human readers.

- We can also define some other basic arithmetic properties of the integers:

- Definition: We can define the binary operation of subtraction in terms of addition and negation by setting $a-b=a+(-b)$.
  - Notice that this operation is well-defined (i.e., the definition makes sense and there is no ambiguity), because $-b$ is unique as we showed above.

- Definition: We define the order relation $<$ (less than) by saying $a<b$ if and only if $b-a\in\mathbb{Z}_+$. We also define $b>a$ (greater than) to mean the same thing, and likewise write $a\leq b$ to mean $a<b$ or $a=b$, and $a\geq b$ to mean $a>b$ or $a=b$.

- The axioms [N1] and [N2] ensure that these symbols all behave in the way we expect inequality symbols to behave.
- Explicitly, [N1] implies that for any integers a and b, exactly one of $a < b$, $a = b$, or $b < a$ holds, because the integer $b - a$ is either positive, zero, or negative (respectively).
- Also, [N2] implies that for any a, b, c with $a < b$ and $b < c$, then $a < c$, because if $b - a$ and $c - b$ are positive, then their sum $(c - b) + (b - a) = c - a$ is also positive.
- Finally, [N2] also implies that for any a, b, c with $a < b$ and $0 < c$, then $ac < bc$, since $b - a$ and $c - 0 = c$ are positive and thus have positive product.

- A seemingly obvious, yet bizarrely important, property of the integers is the following result:
- Proposition: There are no integers between 0 and 1.
  - Observe that this proposition must rely on the well-ordering axiom, because all of the other axioms also apply to the rational and real numbers (which certainly do have elements between 0 and 1).
  - Proof: Let $S = \{r \in \mathbb{Z} : 0 < r < 1\}$ be the set of all integers between 0 and 1. If S is empty, we are done, so assume $S \neq \emptyset$.
  - By the well-ordering axiom [N3], S has a minimal element r.
  - Now observe that since $0 < r < 1$, we have $0 < r^2 < r < 1$ by appropriate uses of [N1] and [N2].
  - But this is a contradiction, because r^2 is then a positive integer less than r, but r was assumed to be minimal.
  - Therefore, S cannot be nonempty, so $S = \emptyset$ as claimed.

2.2 Induction

- We now discuss an important proof technique, called "proof by mathematical induction", that will allow us to prove propositions about all of the positive integers.

2.2.1 Mathematical Induction

- First, we use the well-ordering axiom to establish a fundamental property about sets of positive integers:
- 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 $S = \mathbb{Z}_+$ is the set of all positive integers.
  - Proof: Let $T = \mathbb{Z}_+ \backslash S$, the set of elements of $\mathbb{Z}_+$ not in S. If T is empty, we are done, so assume $T \neq \emptyset$.
  - By the well-ordering axiom [N3], 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, we cannot have $0 < r < 1$.
  - Furthermore, since $1 \in S$, we cannot have $r = 1$.
  - 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 then the hypotheses on S then imply $r \in S$, which is a contradiction since we assumed $r \in T$.
  - Hence $T = \emptyset$, so $S = \mathbb{Z}_+$ as claimed.

- Now we can invoke the result of the proposition to give a concrete procedure for mathematical induction.
  - Explicitly, suppose $P(n)$ is a proposition such that the "base case" $P(1)$ holds, and also such that the "inductive step" holds: namely, $P(n)$ implies $P(n + 1)$ for all $n \geq 1$.
  - Then we claim that $P(k)$ is true for every positive integer k.
  - To show this fact, let S be the set of positive integers k such that $P(k)$ is true.

- By hypothesis, $1 \in S$, and $n \in S$ implies $(n+1) \in S$.
- Therefore, by our proposition, we conclude that S is the set of all positive integers, which is to say, $P(k)$ is true for all positive integers k.

- 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 the 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.

- For example, suppose we wish to show that $1+2+3+4+\cdots+n = \frac{1}{2}n(n+1)$ for every positive integer n.
  - Some quick numerical experimentation will suggest that this formula is correct, but is not likely to suggest a proof.

- To prove that $1+2+3+4+\cdots+n = \frac{1}{2}n(n+1)$ for every positive integer n, we can use the principle of mathematical induction.
  - If we take $P(n)$ to be the statement "$1+2+3+4+\cdots+n = \frac{1}{2}n(n+1)$", then by the inductive principle, all we need to do is show that $P(1)$ is true and that $P(n)$ implies $P(n+1)$ for each $n \geq 1$.
  - The statement $P(1)$ simply reads $1 = \frac{1}{2} \cdot 1 \cdot 2$, which is clearly true.
  - The statement $P(n)$ says that $1+2+3+4+\cdots+n = \frac{1}{2}n(n+1)$, while the statement $P(n+1)$ says that $1+2+3+4+\cdots+n+(n+1) = \frac{1}{2}(n+1)(n+2)$.
  - To prove that $P(n)$ implies $P(n+1)$, we need to start from the statement $1+2+3+4+\cdots+n = \frac{1}{2}n(n+1)$ and use it (somehow) to show that $1+2+3+4+\cdots+n+(n+1) = \frac{1}{2}(n+1)(n+2)$.
  - We can do this as follows: observe that

$$\begin{aligned} 1+2+3+4+\cdots+n+(n+1) &= [1+2+3+4+\cdots+n]+(n+1) \\ &= \frac{1}{2}n(n+1)+(n+1) \\ &= \frac{1}{2}n^2+\frac{1}{2}n+n+1 \\ &= \frac{1}{2}(n^2+3n+2) = \frac{1}{2}(n+1)(n+2). \end{aligned}$$

    where we applied the "inductive hypothesis" piece of information that $1+2+3+4+\cdots+n = \frac{1}{2}n(n+1)$ to go from the first line to the second, and then simply did algebra to rearrange the result into the desired expression.
  - Since we have proven the two required pieces, namely that $P(1)$ is true and that $P(n)$ implies $P(n+1)$, by the principle of mathematical induction, $P(k)$ is true for every $k \geq 1$.

- Induction arguments are useful because they can convert difficult direct proofs into (often) comparatively routine exercises.
  - The base case is usually an easy example where the result is obvious or almost obvious, while the inductive step gives a clear hypothesis to start with and an equally clear goal to reach.
  - Generally, most of the work in the proof goes into the proof of the inductive step.

2.2.2 Examples of Induction Arguments

- Here are a few more examples of proofs by induction, written in a more typical style.
- Example: Prove that $2^0 + 2^1 + 2^2 + \cdots + 2^{n-1} = 2^n - 1$ for every positive integer n.
  - We prove this by induction on n.
  - For the base case $n = 1$, we must show that $2^0 = 2^1 - 1$ which is clearly true.
  - For the inductive step, we are given that $2^0 + 2^1 + 2^2 + \cdots + 2^{n-1} = 2^n - 1$ and must show that $2^0 + 2^1 + 2^2 + \cdots + 2^n = 2^{n+1} - 1$.
  - By the inductive hypothesis, we can write

$$\begin{aligned} 2^0 + 2^1 + 2^2 + \cdots + 2^n &= [2^0 + 2^1 + 2^2 + \cdots + 2^{n-1}] + 2^n \\ &= [2^n - 1] + 2^n \\ &= 2^{n+1} - 1 \end{aligned}$$

    and therefore we see $2^0 + 2^1 + 2^2 + \cdots + 2^n = 2^{n+1} - 1$, as required.
  - By induction, $2^0 + 2^1 + 2^2 + \cdots + 2^{n-1} = 2^n - 1$ for every positive integer n.
- 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.
- Example: Prove that $\dfrac{1}{1\cdot 2} + \dfrac{1}{2\cdot 3} + \cdots + \dfrac{1}{n\cdot(n+1)} = \dfrac{n}{n+1}$ for every positive integer n.
  - We prove this by induction on n.
  - For the base case $n = 1$, we must show that $\dfrac{1}{1\cdot 2} = \dfrac{1}{2}$ which is clearly true.
  - For the inductive step, we are given that $\dfrac{1}{1\cdot 2} + \dfrac{1}{2\cdot 3} + \cdots + \dfrac{1}{n\cdot(n+1)} = \dfrac{n}{n+1}$ and must show that $\dfrac{1}{1\cdot 2} + \dfrac{1}{2\cdot 3} + \cdots + \dfrac{1}{n\cdot(n+1)} + \dfrac{1}{(n+1)\cdot(n+2)} = \dfrac{n+1}{n+2}$.
  - By the inductive hypothesis, we can write

$$\begin{aligned} \frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \cdots + \frac{1}{n\cdot(n+1)} + \frac{1}{(n+1)\cdot(n+2)} &= \left[\frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \cdots + \frac{1}{n\cdot(n+1)}\right] + \frac{1}{(n+1)\cdot(n+2)} \\ &= \frac{n}{n+1} + \frac{1}{(n+1)\cdot(n+2)} \\ &= \frac{n(n+2)+1}{(n+1)(n+2)} = \frac{(n+1)^2}{(n+1)(n+2)} = \frac{n+1}{n+2} \end{aligned}$$

    and therefore we see $\dfrac{1}{1\cdot 2} + \dfrac{1}{2\cdot 3} + \cdots + \dfrac{1}{n\cdot(n+1)} + \dfrac{1}{(n+1)\cdot(n+2)} = \dfrac{n+1}{n+2}$, as required.
  - By induction, $\dfrac{1}{1\cdot 2} + \dfrac{1}{2\cdot 3} + \cdots + \dfrac{1}{n\cdot(n+1)} = \dfrac{n}{n+1}$ 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.
  - We often want to start at a different base case than $n = 1$: frequently, we instead start at $n = 0$ or $n = 2$.
  - As long as we establish the appropriate base case and inductive step, the inductive principle will still work.
  - If, for example, our base case is $n = 2$, then we would prove $P(2)$ is true and that $P(n)$ implies $P(n+1)$, with the conclusion being that $P(k)$ is true for all integers $k \geq 2$.
- Example: Show that $2^n > n^2$ for all integers $n \geq 5$.
  - We prove this by induction on n.
  - For the base case $n = 5$, we must show that $2^5 > 5^2$, or $32 > 25$, which is clearly 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 $n \geq 5$, $2n^2 = n^2 + n^2 \geq n^2 + 5n \geq n^2 + 2n + 1 = (n+1)^2$.
  - 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 $P(n+1)$.
  - It may seem like we are assuming extra information, but in fact strong induction and regular induction are logically equivalent.
  - The reason is that we can view any strong-induction proof as a regular-induction proof with a slightly different hypothesis.
  - Explicitly, if we define $Q(n)$ to be the proposition $Q(n) = [\forall k \in \mathbb{Z}_+, (k \leq n) \Rightarrow P(k)]$, which is to say $Q(n)$ is true whenever $P(1)$, $P(2)$, ..., and $P(n)$ are all true, then a strong-induction proof of the proposition $P(n)$ is the same as a standard-induction proof of the proposition $Q(n)$.
  - Thus, it is always allowable to assume the strong induction hypothesis when writing an induction proof (although in practice, one typically only does so when it is actually necessary).
- Example: Prove that every positive integer can be written as the sum of one or more distinct powers of 2.
  - We will show this by (strong) induction on the integer, n.
  - We take the base case $n = 1$: clearly, $n = 2^0 = 1$ has the required property, as claimed.
  - For the inductive step, suppose that $n \geq 2$ and the result holds for any positive integer less than n.
  - If n is even, then $n/2$ is a positive integer with $n/2 < n$, so by the inductive hypothesis, $n/2$ can be written as the sum of one or more distinct powers of 2, say, $n/2 = 2^{a_1} + \cdots + 2^{a_d}$.
  - Then doubling all of the terms in this sum yields $n = 2^{a_1+1} + \cdots + 2^{a_d+1}$ so n is also the sum of distinct powers of 2, as required.
  - If n is odd, then $(n-1)/2$ is a positive integer with $(n-1)/2 < n$, so by the inductive hypothesis, $(n-1)/2$ can be written as the sum of one or more distinct powers of 2, say, $(n-1)/2 = 2^{a_1} + \cdots + 2^{a_d}$.
  - Then doubling all of the terms and adding 1 yields $n = 2^0 + 2^{a_1+1} + \cdots + 2^{a_d+1}$ so n is also the sum of distinct powers of 2, as required.
- Example: A chocolate bar consists of $A = mn$ identical squares of chocolate arranged in an $m \times n$ rectangular grid. You may break any piece along any row or column along the lines separating the squares to create two separate pieces. By repeatedly breaking pieces, one at a time, prove that the minimum number of breaks required to separate the bar into mn separate 1×1 squares is $A - 1$.
  - We will show this by (strong) induction on the area A.

- We take the base case $A = 1$, in which case $m = n = 1$. In this case, the bar is already separated into 1×1 squares, and we have indeed used $mn - 1 = 0$ breaks as claimed.
- For the inductive step, suppose that the result holds for any bar of total area at most $A-1$, and suppose we have a bar with area A.
- Since rows and columns are interchangeable, assume without loss of generality that we break the bar along a column. This will produce two pieces of chocolate of sizes $m \times n_1$ and $m \times n_2$ for some positive integers n_1 and n_2 with $n_1 + n_2 = n$.
- Since both of these pieces have area less than A (since the sum of their areas is A and they both have positive area), the inductive hypothesis applies to both of them. Therefore, we see that breaking the $m \times n_1$ piece of chocolate into individual squares requires $mn_1 - 1$ breaks, and breaking the $m \times n_2$ piece of chocolate into individual squares requires $mn_2 - 1$ breaks.
- Therefore, the total number of breaks for the original $m \times n$ piece of chocolate is $1 + (mn_1 - 1) + (mn_2 - 1) = m(n_1 + n_2) - 1 = mn - 1 = A - 1$, as claimed. This establishes the inductive step.
- Hence by (strong) induction, the result holds for every positive integer area A.

- In some situations, it may be necessary to have multiple base cases depending on the structure of the induction argument.
  - To illustrate how this can happen, suppose that we are trying to prove $P(n)$ for all positive integers n, and the proof of the inductive step $P(n)$ requires both $P(n-1)$ and $P(n-2)$ to be true.
  - Then it is not sufficient to start with the base case $P(1)$, because using the inductive step to establish $P(3)$ requires that $P(1)$ and $P(2)$ both be known to be true.
  - On the other hand, if we do show both $P(1)$ and $P(2)$ are true, then the inductive step would tell us that $P(3)$ is also true.
  - Then because we know $P(2)$ and $P(3)$, the inductive step would tell us that $P(4)$ is also true, and so on and so forth.

- Example: Let $a_0 = 2$, $a_1 = 5$, and, for $n \geq 2$, let $a_n = 5a_{n-1} - 6a_{n-2}$. Prove that $a_n = 2^n + 3^n$ for all $n \geq 0$.
  - We will show this by strong induction on n.
  - For $n = 0$ and $n = 1$, the result is obvious, since $a_0 = 2^0 + 3^0$ and $a_1 = 2^1 + 3^1$.
  - Now suppose $n \geq 2$. By the strong induction hypothesis and the fact that $n \geq 2$, we have $a_{n-1} = 2^{n-1} + 3^{n-1}$ and $a_{n-2} = 2^{n-2} + 3^{n-2}$, and we want to show that $a_n = 2^n + 3^n$.
  - By the recursion and the induction hypotheses,

$$\begin{aligned} a_n &= 5a_{n-1} - 6a_{n-2} \\ &= 5(2^{n-1} + 3^{n-1}) - 6(2^{n-2} + 3^{n-2}) \\ &= 4 \cdot 2^{n-2} + 9 \cdot 3^{n-2} = 2^n + 3^n \end{aligned}$$

    and therefore $a_n = 2^n + 3^n$ as claimed.
  - By (strong) induction, we conclude that $a_n = 2^n + 3^n$ for all integers $n \geq 0$.

- In some cases the exact nature of the cases being used in the inductive step can be non-obvious, so it is important to be very careful, as the following (famously incorrect) argument shows:

- Incorrect Proposition: All horses are the same color.
  - Proof: We show this result by strong induction on n, the number of horses. The base case $n = 1$ is obvious, since any one horse is the same color as itself.
  - For the inductive step, suppose it is known that any $n - 1$ horses are the same color, and we are given n horses.
  - Then the first $n - 1$ horses are the same color by the inductive hypothesis, and the last $n - 1$ horses are also the same color also by the inductive hypothesis.

- Therefore, every horse is the same color as the middle $n-2$ horses, as required, so all n horses are the same color.
- Hence by induction, the result holds for every positive integer n.
- Remark: Of course, the result is false, but the mistake is reasonably well-hidden: in the proof of the inductive step, it is implicitly assumed that $n-2$ is positive so that $n \geq 3$, but since only the base case $n = 1$ was actually established, the proof is missing an argument for what happens with $n = 2$. (Of course, the result is not true for $n = 2$!)

- As a final remark, we note that it is also possible to phrase induction arguments as "smallest counterexample" or "infinite descent" arguments.
  - The general idea is to work to show that $P(n)$ is true for all positive integers n by contradiction.
  - If $P(n)$ is not true for all positive integers n, then by the well-ordering axiom there must exist a minimal positive integer k such that $P(k)$ is false: this would be a "minimal counterexample".
  - If one can then prove that the existence of such a counterexample would imply the existence of a smaller counterexample (i.e., some smaller positive integer k' such that $P(k')$ is false), this would yield a contradiction.
  - Notice that the structure of this argument is equivalent to proof by induction, since both arguments invoke the well-ordering axiom as a way of showing that a set is equal to $\mathbb{Z}_+$.
  - In certain cases it can be easier to identify salient features of the induction argument by phrasing the problem in terms of smallest counterexamples. However, standard proof by induction tends to be more straightforward since it is a direct proof rather than a proof by contradiction.

- Example: Prove that every positive integer can be written as the sum of one or more distinct powers of 2.
  - Suppose otherwise, so that there is at least one positive integer that cannot be written as the sum of one or more distinct powers of 2, and choose the smallest such integer n.
  - Then because n is minimal and clearly $n > 1$, this means $n-1$ can be written as the sum of one or more distinct powers of 2. If all of these terms were even, then we could simply add 1 to the sum to obtain a representation for n, which contradicts the assumption that n is a counterexample.
  - This means that $n-1$ must have a term of 1 in its sum, and so n is even. But now consider $n/2$ instead: it is a positive integer less than n, so again by minimality we would have such a representation for $n/2$. But doubling all of the terms would yield a representation for n, which is again a contradiction.
  - We obtain a contradiction in both cases, so there cannot exist any such n.

2.3 Divisibility and the Euclidean Algorithm

- We have constructed three of the operations of standard arithmetic: $+$, $-$, and $\cdot$. We now discuss division.
  - One caveat with division is that, unlike addition, subtraction, and multiplication, the quotient of one integer by another (even if it is defined) need not be an integer.
  - Thus, instead of discussing division, we start by discussing divisibility.

2.3.1 Divisibility and Division With Remainder

- Definition: If $a \neq 0$, we say that a divides b, written $a|b$, if there exists an integer k with $b = ka$. If $a|b$, we also say that b is divisible by a.
  - Examples: $2|4$ since $4 = 2 \cdot 2$, $(-7)|7$ since $7 = (-1) \cdot (-7)$, $13|1001$ since $1001 = 77 \cdot 13$, $6|0$ since $0 = 0 \cdot 6$, and $0|0$ since $0 = 2019 \cdot 0$.
  - If a does not divide b, we sometimes write $a \nmid b$. For example, $2 \nmid 3$ since there is no integer k with $3 = 2k$.
  - In the particular case of divisibility by 2, we say n is even if $2|n$. We will show (carefully) later that $2 \nmid n$ is equivalent to saying that $2|(n-1)$, which we take as the definition of odd.

- There are a number of basic properties of divisibility that follow from the definition and properties of arithmetic:
- Proposition (Properties of Divisibility): For any integers a, b, c, m, x, y, the following hold:
  1. If $a|b$, then $a|bc$ for any c.
  2. If $a|b$ and $b|c$, then $a|c$.
  3. If $a|b$ and $a|c$, then $a|(xb + yc)$ for any x and y.
  4. If $a|b$ and $b|a$, then $a = \pm b$.
  5. If $a|b$, and $a, b > 0$, then $a \leq b$.
  6. For any $m \neq 0$, $a|b$ is equivalent to $(ma)|(mb)$.
     - Proof: Each of these follows essentially directly from the definition of divisibility and the basic properties of arithmetic.
     - For example, (2) follows because $a|b$ and $b|c$ imply that there exist integers k and l such that $b = ka$ and $c = lb$, and thus $c = lb = (lk)a$: hence c is an integer times a, so $a|c$.
     - Likewise, (5) follows because if $a|b$ and a, b are positive, then $b = ka$ for some positive integer k. Since this means $1 \leq k$ because there are no integers between 0 and 1, we have $a \leq ka = b$, and so $a \leq b$.
- 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 With Remainder): 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 second statement follows immediately from the first statement: if $r = 0$ then $a = qb$ so $b|a$, and if $b|a$ then $a = kb$ for some k; then the uniqueness of q and r implies that we must have $q = k$ and $r = 0$.
  - To show existence of q and r, 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 $r < b$. But since r is in the set S, we must have $r = a - qb$ for some integer q. Therefore, $a = qb + r$ for some integers q, r such that $0 \leq r < b$, as required.
  - 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 between -1 and 0), it must be the case that $q' - q = 0$, so that $q' = q$ and then $r' = r$.
  - Example: If $a = 25$ and $b = 4$, then the set $S = \{\ldots, -7, -3, 1, 5, 9, 13, 17, 21, 25, 29, 33, \ldots\}$, and $T = \{1, 5, 9, 13, 17, 21, 25, 29, 33, \ldots\}$. 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.
- As an immediate consequence of the existence of the quotient and remainder in the division algorithm, we see that every integer is either even (i.e., leaves a remainder of 0 when divided by 2) or odd (i.e., leaves a remainder of 1 when divided by 2), and the uniqueness of the quotient and remainder imply that no integer is both even and odd.

2.3.2 Greatest Common Divisors and Least Common Multiples

- We now discuss the idea of common divisors.
- 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 greatest 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 notation used in abstract algebra. We will always 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 so $\gcd(30, 42) = 6$.

- Our first main result about greatest common divisors is that we can write them in terms of the original integers:

- Theorem (GCD as Linear Combination): If a and b are integers, not both zero, and $d = \gcd(a, b)$, then there exist integers x and y with $d = ax + by$: in fact, the gcd is the smallest positive such linear combination.
  - 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 $S = \{as + bt : s, t \in \mathbb{Z}\} \cap \mathbb{Z}_+$.
  - Clearly $S \neq \emptyset$ since one of a and $-a$ is in S, so now let $l = ax + by$ be the minimal element of S.
  - We claim that $l|b$. To see this, apply the division algorithm to write $b = ql + r$ for some $0 \leq r < l$.
  - Then $r = b - ql = b - q(ax + by) = a(-qx) + b(1 - qy)$ 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.
  - This leaves only the possibility $r = 0$, and therefore we have $l|b$.
  - By a symmetric argument, $l|a$, and so l is a common divisor of a and b. This requires $l \leq d$.
  - But now since $d|a$ and $d|b$ we can write $a = dk_a$ and $b = dk_b$ for some integers k_a and k_b. Then $l = ax + by = dk_a x + dk_b y = d(k_a x + k_b y)$.
  - Therefore we see also that $d|l$, so in particular $d \leq l$ since both are positive. Since $l \leq d$ as well from above, we must have $l = d$.

- Corollary: If $l|a$ and $l|b$, then l divides $\gcd(a, b)$. In other words, the gcd of a and b is divisible by every other common divisor.
  - 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.
  - Examples: 24 and 13 are relatively prime. 2 and 5 are relatively prime. 15 and 16 are relatively prime.
  - Non-Example: 30 and 69 are not relatively prime, since they have the common divisor 3.

- Using all of the results we have shown above, we can collect a number of useful facts about greatest common divisors:

- Proposition (Properties of GCDs): If m, a, b, d are integers, then the following hold:

1. If $m > 0$, then $\gcd(ma, mb) = m \cdot \gcd(a, b)$.
   - Proof: As shown above, $\gcd(ma, mb)$ is the smallest positive element of the set $S = \{max + mby : x, y \in \mathbb{Z}\}$, while $\gcd(a, b)$ is the smallest positive element of the set $T = \{ax + by : x, y \in \mathbb{Z}\}$.
   - But clearly, multiplying all of the elements of T by m yields the set S, and since this operation preserves the identity of the smallest positive element, we must have $\gcd(ma, mb) = m \cdot \gcd(a, b)$, as claimed.
2. If $d > 0$ divides both a and b, then $\gcd(a/d, b/d) = \gcd(a, b)/d$.

- Proof: Applying (1) to a/d and b/d with $m = d$ yields $\gcd(a, b) = d \cdot \gcd(a/d, b/d)$, and then dividing both sides by d yields the required statement.

3. There exist integers x and y with $ax + by = 1$ if and only if $\gcd(a, b) = 1$.
   - Proof: If $\gcd(a, b) = 1$ then we showed above that there exist integers x and y with $ax + by = 1$.
   - For the other direction, any common divisor of a and b must divide $ax + by = 1$: hence the gcd must divide 1, which leaves only the possibility that it equals 1.
4. If a and b are both relatively prime to m, then so is ab.
   - Proof: By the linear combination property of the gcd, there exist x_1, y_1, x_2, y_2 with $ax_1 + my_1 = 1$ and $bx_2 + my_2 = 1$.
   - Multiplying these two equations together and rearranging the results yields $ab(x_1x_2) + m(y_1bx_2 + y_2ax_1 + my_1y_2) = 1$, and this implies that ab and m are relatively prime.
5. For any integer x, $\gcd(a, b) = \gcd(a, b + ax)$.
   - Proof: Observe that the set of linear combinations of a and b is the same as the set of integral linear combinations of a and $b + ax$.
6. If $a|bc$ and a and b are are relatively prime, then $a|c$.
   - Proof 1: By (1), we have $\gcd(ac, bc) = c \cdot \gcd(a, b) = c$. Since $a|bc$ and $a|ac$, we see that a is a common divisor of ac and bc, and therefore divides the gcd, which is c. Thus $a|c$ as claimed.
   - Proof 2: Since a and b are relatively prime, by (3) there exist integers x and y with $ax + by = 1$. Multiplying both sides by c yields $acx + bcy = c$: but now note that a divides both acx and bcy, so a must also divide their sum c.

- Dual to the notion of the greatest common divisor is the notion of the least common multiple:
- 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, as follows by noting that $210 = 7 \cdot 30 = 5 \cdot 42$ and that none of $1 \cdot 42$, $2 \cdot 42$, $3 \cdot 42$, and $4 \cdot 42$ is divisible by 30.
  - 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, but it turns out that we can obtain either one from the other:
- Proposition (Properties of LCMs): If m, a, b are any positive integers, then the following hold:

1. We have $\text{lcm}(ma, mb) = m \cdot \text{lcm}(a, b)$.
   - Proof: 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. Thus $k \geq l$, where $l = \text{lcm}(a, b)$.
   - On the other hand, certainly ma and mb divide ml, so $ml \geq mk$. We must therefore have $l = k$, so $\text{lcm}(ma, mb) = m \cdot \text{lcm}(a, b)$ as claimed.
2. If a and b are positive integers, then $\gcd(a, b) \cdot \text{lcm}(a, b) = ab$.
   - 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 (1) 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.

2.3.3 The Euclidean Algorithm

- Although we have identified various properties of the gcd, we have not yet described a convenient procedure for actually computing the gcd other than by writing down lists of common divisors. (Nor have we described how to compute the gcd as a linear combination of the original integers.) Both questions turn out to 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:

$$\begin{aligned} a &= q_1 b + r_1 \\ b &= q_2 r_1 + r_2 \\ r_1 &= q_3 r_2 + r_3 \\ &\vdots \\ r_{k-1} &= q_{k+1} r_k + r_{k+1} \\ r_k &= q_{k+2} r_{k+1}. \end{aligned}$$

  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 axiom dictates that we cannot have 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 by an easy induction, we claim that $\gcd(r_j, r_{j+1}) = \gcd(a, b)$ for each $0 \leq j \leq k$, where we set $r_0 = b$ and $r_{-1} = a$. The base case $j = 0$ follows from $\gcd(a, b) = \gcd(b, r_1)$ above, and the inductive step follows by applying the same argument to see $\gcd(r_j, r_{j+1}) = \gcd(r_{j+1}, r_{j+2})$.
  - We conclude that $\gcd(a, b) = \gcd(r_{k+1}, r_k) = r_{k+1}$ since r_{k+1} divides r_k. Hence, $\gcd(a, b)$ is the last nonzero remainder as claimed.
  - The correctness of the algorithm for computing the gcd also follows by an easy induction: explicitly, we show by induction on j that there exist integers x_j and y_j such that $r_j = x_j a + y_j b$ for all integers j with $0 \leq j \leq k + 1$.
  - The base cases $j = 0$ and $j = 1$ follow by writing $r_0 = b$ and $r_1 = a - q_1 b$ so we may take $x_0 = 0$, $y_0 = 1$, $x_1 = 1$, and $y_1 = -q_1$.
  - The inductive step follows by writing $r_{j-1} = q_{j+1} r_j + r_{j+1}$, so rearranging yields $r_{j+1} = r_{j-1} - q_{j+1} r_j = (x_{j-1} a + y_{j-1} b) - q_{j+1}(x_j a + y_j b) = (x_{j-1} - q_{j+1} x_j) a + (y_{j-1} - q_{j+1} y_j) b$ and thus we take $x_{j+1} = x_{j-1} - q_{j+1} x_j$ and $y_{j+1} = y_{j-1} - q_{j+1} y_j$.
  - By induction, we eventually obtain an expression $\gcd(a, b) = r_{k+1} = x_{k+1} a + y_{k+1} b$ as required.

- Example: Find the gcd of 133 and 98 using the Euclidean algorithm, and write the gcd explicitly as a linear combination of 133 and 98.

  - First, we use the Euclidean algorithm:

$$\begin{aligned} 133 &= 1 \cdot 98 + 35 \\ 98 &= 2 \cdot 35 + 28 \\ 35 &= 1 \cdot 28 + 7 \\ 28 &= 4 \cdot 7 \end{aligned}$$

    and so the gcd is $\boxed{7}$.

- For the linear combination, we solve for the remainders:

$$\begin{array}{rcccccl} 35 & = & 133 - 1 \cdot 98 & = & & = & 1 \cdot 133 - 1 \cdot 98 \\ 28 & = & 98 - 2 \cdot 35 & = & 98 - 2 \cdot (133 - 1 \cdot 98) & = & -2 \cdot 133 + 3 \cdot 98 \\ 7 & = & 35 - 1 \cdot 28 & = & (1 \cdot 133 - 1 \cdot 98) - 1 \cdot (-2 \cdot 133 + 3 \cdot 98) & = & 3 \cdot 133 - 4 \cdot 98 \end{array}$$

  so we obtain $\boxed{7 = 3 \cdot 133 - 4 \cdot 98}$.

- 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 44773 and 2088 using the Euclidean algorithm, and use the results to write the gcd as an explicit linear combination.

  - Applying the Euclidean algorithm to $a = 44773$ and $b = 2088$ yields

$$\begin{aligned} 44773 &= 5 \cdot 8537 + 2088 \\ 8537 &= 4 \cdot 2088 + 185 \\ 2088 &= 11 \cdot 185 + 53 \\ 185 &= 3 \cdot 53 + 26 \\ 53 &= 2 \cdot 26 + 1 \\ 26 &= 26 \cdot 1 \end{aligned}$$

  - For the linear combination, we solve for the remainders:

$$\begin{array}{rcccl} 2088 & = & & = & 1 \cdot 44773 - 5 \cdot 8537 \\ 185 & = & 8537 - 4 \cdot 2088 & = & -4 \cdot 44773 + 21 \cdot 8537 \\ 53 & = & 2088 - 11 \cdot 185 & = & 45 \cdot 44773 - 236 \cdot 8537 \\ 26 & = & 185 - 3 \cdot 53 & = & -139 \cdot 44773 + 729 \cdot 8537 \\ 1 & = & 53 - 2 \cdot 26 & = & 323 \cdot 44773 - 1694 \cdot 8537 \end{array}$$

  and therefore we can take $s = \boxed{323}$ and $t = \boxed{-1694}$.

2.4 Primes and Unique Factorization

- Now that we have examined divisibility and common factors, we will examine one of the other fundamental properties of the integers, namely, the existence and uniqueness of prime factorizations.

- We begin by discussing prime numbers:

- Definition: If $p > 1$ is an integer, we say it is prime if there is no integer d with $1 < d < p$ such that $d|p$. (In other words, p is prime if p has no proper divisors.) If $n > 1$ is not prime, which is to say, if there exists some integer d with $1 < d < n$ with $d|n$, we say it is composite.

  - The first few primes are 2, 3, 5, 7, 11, 13, 17, 19, and so forth. 1 is neither prime nor composite.
  - Remark: In more advanced contexts, the following equivalent definition of a prime is often used instead: the integer $p > 1$ is prime if and only if $p|ab$ implies that $p|a$ or $p|b$.

- The prime numbers are often called the "building blocks under multiplication", because every positive integer can be written as the product of prime numbers in an essentially unique way. To prove this, we first show that there exists at least one such factorization:

- Proposition (Existence of Prime Factorizations): Every positive integer n can be written as a product of zero or more primes (where a "product" is allowed to have only one term, and the empty product has value 1).

  - The representation of n as a product of primes is called the prime factorization of n. (For example, the prime factorization of 6 is $6 = 2 \cdot 3$.) We will show in a moment that it is unique up to reordering the terms.

- Proof: We use strong induction on n. The result clearly holds if $n = 1$, since 1 is the empty product.
- Now suppose $n \geq 2$. If n is prime, we are done (simply take the product n with one term), so assume that n is not prime, hence 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.

- To establish the uniqueness of prime factorizations, we require the following prime divisibility property:
- Proposition (Prime Divisibility): If a and b are integers and p is a prime number with $p|ab$, then $p|a$ or $p|b$.
  - Proof: 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 since p is prime. But the gcd cannot be 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 $p|b$.
- Theorem (Fundamental Theorem of Arithmetic): Every positive integer can be factored into a product of primes, and this factorization is unique up to reordering of the factors.
  - Proof: We already showed that every positive integer has a prime factorization, so we need only show the uniqueness.
  - Suppose by way of contradiction that there is a positive integer with two prime factorizations. By the well-ordering axiom, we may select the minimal such positive integer n with 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, so 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 often write the prime factorization in the 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[3].
- Proposition (Divisibility and Factorizations): If $a = \prod_{i=1}^{j} p_i^{a_i}$ and $b = \prod_{i=1}^{j} p_i^{b_i}$ for distinct primes p_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 $\mathrm{lcm}(a, b) = \prod_{i=1}^{j} p_i^{\max(a_i, b_i)}$.
  - 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 follow directly: the exponent of p_i in the gcd is the largest integer that is $\leq a_i$ and $\leq b_i$, which is simply the minimum of a_i and b_i, and the exponent of p_i in the lcm is the least integer that is $\geq a_i$ and $\geq b_i$, which is simply the maximum of a_i and b_i.
- Example: For $a = 2^3 3^{10} 5^4$ and $b = 2^4 3^3 5^4 7$, we have $\gcd(a, b) = 2^3 3^3 5^4$ and $\mathrm{lcm}(a, b) = 2^4 3^{10} 5^4 7$.
- 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).

[3]The notation $\prod_{i=1}^{j} f(i)$ is shorthand for the product $f(1)f(2)\cdots f(j)$, in the same way that the notation $\sum_{i=1}^{j} f(i)$ is shorthand for the sum $f(1) + f(2) + \cdots + f(j)$.

- 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.

- At this stage, we will briefly mention a few of the most famous results and open problems relating to prime numbers:

  - (Prime Number Theorem) Euclid's result, while extremely elegant, does not tell us much about the actual primes themselves: for example, it does not say anything about how common the primes are. Are most numbers prime? Or are most numbers composite? A more rigorous way to frame this question is: let $\pi(n)$ be the number of primes in the interval $[1, n]$. How fast does $\pi(n)$ increase as n increases: does it grow like n, or $\sqrt{n}$, or something else? The answer is given by the so-called "Prime Number Theorem": $\pi(n) \sim \dfrac{n}{\log(n)}$, where log denotes the natural logarithm. (The notation $f(n) \sim g(n)$ means that as $n \to \infty$, the limit $\lim_{n\to\infty} f(n)/g(n) = 1$.)
  - (Twin Primes) Another question is: how close do primes get? It is obvious that 2 is the only even prime, so aside from 2 and 3, any pair of primes has to differ by at least 2: such pairs are called "twin primes". One can write down a long list of twin primes: (3,5), (5,7), (11,13), (17,19), (29,31), (41,43), (59,61), and so forth. Are there infinitely many? The answer is not known, although twin primes are expected to be quite rare. However, it has been proven (as of August 2014) that there exist infinitely many pairs of primes (p_1, p_2) such that $|p_2 - p_1| \leq 246$.
  - (Goldbach's Conjecture) One can observe that $2 + 2 = 4$, $3 + 3 = 6$, $3 + 5 = 8$, $3 + 7 = 10$, $5 + 7 = 12$, $3 + 11 = 14$, $3 + 13 = 16$, $5 + 13 = 18$, $7 + 13 = 20$, and so forth. It appears that every even number (bigger than 4) can be written as the sum of two primes. It is not known whether this pattern continues, although it has been numerically verified for every even integer less than 10^{18}. In 2013, a proof that every odd integer greater than 7 can be written as a sum of three primes was announced. (This result is weaker than Goldbach's conjecture, but it is of the same type.)

- There are many applications of prime factorizations in elementary number theory, but one particularly famous result is that $\sqrt{2}$ is irrational:

- Theorem (Irrationality of $\sqrt{2}$): The number $\sqrt{2}$ is irrational, which is to say, there do not exist integers m and n such that $\sqrt{2} = m/n$.

  - Proof: Suppose by way of contradiction that $\sqrt{2}$ were rational so that $\sqrt{2} = m/n$ for some integers m and n, which (by negating if needed) we may assume are positive.
  - Squaring both sides and clearing denominators yields the equivalent equation $2n^2 = m^2$.
  - Now consider the prime factorizations of both sides: say $m = 2^{m_2} 3^{m_3} \cdots$ and $n = 2^{n_2} 3^{n_3} \cdots$.
  - We obtain the equality $2^{2m_2+1} 3^{m_3} \cdots = 2^{2n_2} 3^{2n_3} \cdots$, and so by the uniqueness of prime factorizations, all of the corresponding exponents must be equal.
  - In particular, we see that $2m_2 + 1 = 2n_2$, so that $2(n_2 - m_2) = 1$. But this is impossible, because 2 does not divide 1.
  - Therefore, it could not have been true that $\sqrt{2} = m/n$, so $\sqrt{2}$ must be irrational as claimed.

2.5 Modular Congruences and The Integers Modulo m

- 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. This is quite natural despite the fact that $3+11$ is 14, not 2: simply put, we identify times that are 12 hours apart as the same time of day.

2.5.1 Modular Congruences and Residue Classes

- Modular arithmetic is simply a formalization of this "clock arithmetic":
- 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 "$a \equiv b \pmod{m}$".
  - 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. $a \equiv a \pmod{m}$.
2. $a \equiv b \pmod{m}$ if and only if $b \equiv a \pmod{m}$.
3. If $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m}$, then $a \equiv c \pmod{m}$.
4. If $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, then $a+c \equiv b+d \pmod{m}$.
5. If $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, then $ac \equiv bd \pmod{m}$.
6. If $a \equiv b \pmod{m}$, then $ac \equiv bc \pmod{mc}$ for any $c > 0$.
7. If $d|m$, then $a \equiv b \pmod{m}$ implies $a \equiv b \pmod{d}$.
   - Proof: Each of these follows in a relatively straightforward way from the definition of modular congruence. The trickiest is (5), which follows by observing that if m divides $b-a$ and m divides $d-c$, then m divides $bd-ac = b(d-c)+a(b-a)$.

- 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 $\{\ldots, -6, -2, 2, 6, 10, 14, \ldots\}$.
  - Example: The residue class of 2 modulo 5 is the set $\{\ldots, -8, -3, 2, 7, 12, 17, \ldots\}$.
  - Example: The residue class of 11 modulo 19 is the set $\{\ldots, -27, -8, 11, 30, 49, 68, , \ldots\}$.
- 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, ... , $m-1$, which is to say, every integer lies in precisely one of the residue classes $\overline{0}$, $\overline{1}$, ... , $\overline{m-1}$.

- If we apply results (2) and (3) from the proposition above when $m = 2$, we obtain (once again) the statement that every integer either leaves a remainder of 0 or 1 when divided by 2: i.e., that every integer is either even or odd, and no integer is both.

2.5.2 The Integers Modulo m, Modular Arithmetic

- 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.
- 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:
- Theorem (Invertible Elements in $\mathbb{Z}/m\mathbb{Z}$): If $m > 0$, then the residue class $\overline{a}$ has a multiplicative inverse in $\mathbb{Z}/m\mathbb{Z}$, meaning that there exists some residue class $\overline{x}$ with $\overline{x} \cdot \overline{a} = \overline{1}$, if and only if a and m are relatively prime.
  - Proof: First suppose that a and m are relatively prime. Then by our analysis of the Euclidean algorithm, there exist integers x and y such that $xa + ym = 1$: then $xa \equiv 1 \pmod{m}$, which is to say $\overline{x} \cdot \overline{a} = \overline{1}$, so that $\overline{a}$ has a multiplicative inverse as claimed.
  - Conversely, suppose $\overline{a}$ were invertible in $\mathbb{Z}/m\mathbb{Z}$ with inverse $\overline{x}$. Then we would have $\overline{x} \cdot \overline{a} = \overline{1}$, or equivalently $xa \equiv 1 \pmod{m}$, and this is in turn equivalent to saying there exists an integer y with $xa + ym = 1$. But then the common divisor d would divide $xa + ym$ hence divide 1, and so a and m are relatively prime.

- The proof above shows that we can find the inverse of an invertible residue class via the Euclidean algorithm.
- Example: Find the multiplicative inverse of $\overline{9}$ in $\mathbb{Z}/11\mathbb{Z}$.
  - Using the Euclidean algorithm, we can obtain $1 = 5 \cdot 11 - 6 \cdot 9$. Reducing both sides modulo 11 yields $\overline{1} = \overline{-6} \cdot \overline{9}$, and since $\overline{-6} = \overline{5}$, this shows that the multiplicative inverse of $\overline{9}$ in $\mathbb{Z}/11\mathbb{Z}$ is $\boxed{\overline{5}}$.

- The case where the modulus is prime is of particular importance:
- Corollary: If p is a prime number, then every nonzero residue class in $\mathbb{Z}/p\mathbb{Z}$ has a multiplicative inverse.
  - Proof: If p is prime, then p is relatively prime to each of 1, 2, ... , $p-1$ and hence all of the nonzero residue classes modulo p are invertible.

- This corollary states that when p is prime, $\mathbb{Z}/p\mathbb{Z}$ has the structure of the algebraic object called a field.
  - To summarize: a field is a set F together with two binary operations of addition ($+$) and multiplication ($\cdot$) both of which are associative and commutative and where $\cdot$ distributes over $+$, that also possesses an additive identity 0 and a multiplicative identity $1 \neq 0$, and where every element has an additive inverse and every nonzero element has a multiplicative inverse.
  - Some familiar examples of fields include $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$.

- Although we will end our discussion of modular arithmetic (and number theory more generally) here, well before the true end of the story, we will remark that modular arithmetic is fundamental in many areas of in mathematics, particularly abstract algebra, algebraic topology, and applied mathematics, as well as outside mathematics.
  - More specifically, modular arithmetic is deeply enmeshed in computer science and modern cryptography, as many computational algorithms employ modular arithmetic, as do most current cryptosystems (e.g., RSA, AES, and elliptic-curve cryptography).
  - Modular arithmetic also arises naturally in chemistry (in the study of molecular symmetries), music theory (in the study of tuning systems), economics and game theory (in the study of fair division problems), and the visual arts (in the study of various artistic designs), among other disciplines.

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