Chapter 4

Divisibility

4.1 The division algorithm

For the next few lectures we will exercise our ability to prove mathematical statements, using the fertile ground of number theory. In the process we will learn new proof techniques and tricks of trade. The number-theoretic concepts and results we will cover will be useful throughout your computer science studies, and, indeed, throughout your involvement with mathematics.

The following result is commonly known as the *division algorithm*, even though it is not an algorithm at all.

Theorem 4.1.1. *If a and b are integers and $b \neq 0$, then there is a unique pair of integers q and r, such that $a = qb + r$ and $0 \leq r < |b|$.*

Proof. We need to prove two things: that there is some such pair q, r (existence) and that this pair is unique (uniqueness).

Let's begin with existence. First we show that there is a pair $q, r \in \mathbb{Z}$ that satisfies $a = qb + r$ for some $r \geq 0$. This is easy after some playing around: Take $q = -|ab|/b$ and $r = a + |ab|$. Since $|b| \geq 1$, it holds that $r \geq 0$. Now we need to show that such $q, r \in \mathbb{Z}$ exist with r in addition being smaller than $|b|$. For this, consider the set S of all $r \in \mathbb{N}$ that satisfy $a = qb + r$ for some $q \in \mathbb{Z}$. We've just shown that S is nonempty, so it must have a smallest element, call it r_0. We have $a = q_0 b + r_0$. If $r_0 < |b|$ we're done. Otherwise, we have $a = (q_0 b + |b|) + (r_0 - |b|)$, which means that $r_0 - |b|$ is a smaller element of S than r_0, leading to a contradiction. This completes the existence proof.

To prove uniqueness, suppose that $a = qb + r = sb + t$, with $0 \leq r, t < |b|$. Thus $(q - s)b + (r - t) = 0$. Since $0 \leq r, t < |b|$, we have $|r - t| < |b|$, hence $|(q - s)b| < |b|$ and $|q - s| < 1$. Since q and s are integers, this implies $q = s$. From this we have $r = t$ and the uniqueness proof is complete. □

Proof tip: When we need to prove that some mathematical object exists and is unique, it is useful to approach in two stages. First prove that at least one such object exists. This can be done either by directly constructing an object and demonstrating

that it fulfills the requirements, or by assuming that no such object exists and reaching a contradiction. Then show that any two such objects must be the same.

The Well-Ordering Principle. In proving the division algorithm, we considered a certain set $S \subseteq \mathbb{N}$ and argued that since it is nonempty, it must have a smallest element. Why is this true? As with induction, we accept this proposition as an axiom. In general, the "well-ordering principle" states that *any nonempty set of natural numbers must have a smallest element.* As you will prove in the homework, the well-ordering principle is equivalent to the principles of induction and strong induction.

4.2 Remainders

A more algorithmic view of Theorem 4.1.1 is as follows: If we divide the equation

$$a = qb + r$$

by b we get

$$\frac{a}{b} = q + \frac{r}{b}.$$

Since $0 \leq r < |b|$, we get that if $b > 0$, then $0 \leq \frac{r}{b} < 1$ and thus $q = \left\lfloor \frac{a}{b} \right\rfloor$, the greatest integer less than or equal to $\frac{a}{b}$. If $b < 0$, then $0 \geq \frac{r}{b} > -1$ and thus $q = \left\lceil \frac{a}{b} \right\rceil$, the least integer greater or equal to $\frac{a}{b}$. This can be used to calculate q, from which we can derive r.

In Theorem 4.1.1, we call q the *quotient* and r the *remainder*. We use the notation $r = a \text{ rem } b$ to denote that r is the remainder when a is divided by b. There is no need for a special notation for quotient, since we can just use $\left\lfloor \frac{a}{b} \right\rfloor$ and $\left\lceil \frac{a}{b} \right\rceil$, depending on the sign of b.

Definition: If a and b are such that $a \text{ rem } b = 0$ we say that a is a *multiple* of b, or that b *divides* a (or is a *divisor* of a). Note that this holds when there exists some integer q, such that $a = qb$. In particular, every integer divides 0, and every integer is a multiple of 1. When b divides a we write $b|a$, and when b does not divide a we write $b \not| a$.

Definition: An integer u is called a *linear combination* of a set of integers $a_1, a_2, \ldots, a_n$ if and only if there exist integer coefficients $c_1, c_2, \ldots, c_n$ that satisfy

$$u = \sum_{i=1}^{n} c_i a_i.$$

Theorem 4.2.1. *Properties of divisibility:*

(a) If $b|a$ and $c|b$ then $c|a$.

(b) If $b|a$ and $a \neq 0$ then $|b| \leq |a|$.

(c) If b divides each of $a_1, a_2, \ldots, a_n$, then b divides all linear combinations of $a_1, a_2, \ldots, a_n$.

(d) $a|b$ and $b|a$ if and only if $a = \pm b$.

Proof. We prove the properties in turn:

(a) Since $b|a$, there exists an integer q, such that $a = qb$. Similarly, there exists an integer r, such that $b = rc$. Thus $a = qb = qrc$. Since qr is an integer, it holds that $c|a$.

(b) Since $b|a$, there exists an integer q, such that $a = qb$. This implies $|a| = |q| \cdot |b|$. Assume for the sake of contradiction that $a \neq 0$ but $|b| > |a|$. Then $|q| \cdot |b| < |b|$. Since $|b| > |a| > 0$, we can divide by $|b|$ to get $|q| < 1$, implying $q = 0$. Thus $a = qb = 0$, which is a contradiction.

(c) Consider a linear combination $u = \sum_{i=1}^{n} c_i a_i$. Since $b|a_i$, there exists an integer q_i, such that $a_i = q_i b$, for all $1 \leq i \leq n$. Thus

$$u = \sum_{i=1}^{n} c_i a_i = \sum_{i=1}^{n} c_i q_i b = b \cdot \sum_{i=1}^{n} c_i q_i.$$

Since $\sum_{i=1}^{n} c_i q_i$ is an integer, we have $b|u$.

(d) For the "if" statement, note that if $a = \pm b$ then $b = qa$ and $a = qb$, for $q = \pm 1$, so $a|b$ and $b|a$. To prove the "only if" statement, assume that $a|b$ and $b|a$. This implies the existence of integers q and r, such that $b = qa$ and $a = rb$. Thus $b = qrb$. If $b = 0$ then $a = 0$ and the claim that $a = \pm b$ holds. Otherwise we can divide by b to get $qr = 1$. Note that in this case $q, r \neq 0$. Part (b) of the theorem implies that $|q| \leq 1$ and $|r| \leq 1$. Thus $q, r = \pm 1$ and the claim that $a = \pm b$ follows.

□

Proof tip: Often we need to prove that a proposition A holds if and only if some other proposition B holds. Such an "if and only if" (sometimes abbreviated as "iff") statement is really composed of two implications, each of which needs to be proved. It is often useful to decouple these and prove them separately. First prove that "If A then B," and then prove that "If B then A." Another strategy is to prove that "If A then B" and "If not A then not B."

4.3 Greatest common divisors

If $d|a$ and $d|b$ then d is a *common divisor* of a and b. For example, 1 is a common divisor of any pair a, b. If a and b are not both 0 then, by Theorem 4.2.1(b), any

common divisor of a and b is not greater than $\max(|a|, |b|)$. Thus the set of common divisors of a and b has a largest element, called the *greatest common divisor* of a and b, or $\gcd(a, b)$. This is the integer d that satisfies the following two criteria:

- $d|a$ and $d|b$.
- If $c|a$ and $c|b$ then $c \leq d$.

Note that when $a = b = 0$, there is no greatest common divisor, since any integer divides 0. When a and b are not both 0, we often want to compute $\gcd(a, b)$ efficiently. Note that the set of divisors of a and $-a$ is the same, and similarly for b and $-b$. Furthermore, if $a = 0$ then $\gcd(a, b) = b$, and if $a = b$ then $\gcd(a, b) = a = b$. Thus it suffices to concentrate on the case $a > b > 0$, without loss of generality.

Since $1 \leq \gcd(a, b) \leq b$, we can just test all integers between 1 and b and choose the largest one that divides both a and b. However, there is a much more efficient way to find greatest common divisors, called Euclid's algorithm. This algorithm, one of the earliest in recorded history, is based on the following lemma.

Lemma 4.3.1. *If $a = qb + r$ then $\gcd(a, b) = \gcd(b, r)$.*

Proof. By Theorem 4.2.1(c), all common divisors of b and r also divide a, since a is a linear combination of b and r. Thus a common divisor of b and r is also a common divisor of a and b. Similarly, since $r = a - qb$, a common divisor of a and b also divides r, so it is a common divisor of b and r. Thus a, b and b, r have the same set of common divisors, and in particular the same greatest common divisor. □

With this lemma in our toolbelt, Euclid's algorithm is easy to describe. To find $\gcd(a, b)$, use the division algorithm (Theorem 4.1.1) to represent $a = qb + r$, where $0 \leq r < b$. (Remember that we are assuming that $a > b > 0$.) If $r = 0$ then $b|a$ and $\gcd(a, b) = b$. Otherwise $\gcd(a, b) = \gcd(b, r)$ and $b > r > 0$. We can thus repeat the above procedure recursively with the pair b, r. Every recursive call strictly reduces both numbers in the pair, so after at most b steps the algorithm will terminate with a valid greatest common divisor of a and b. You will formally prove the correctness of the algorithm in the homework.

4.4 Greatest common divisors and linear combinations

We have seen that a common divisor of a and b divides any linear combination of a and b. Now we will prove a surprising property known as *Bezout's identity* that shows that the greatest common divisor of a and b is itself a linear combination of a and b.

Theorem 4.4.1. *For two integers a and b that are not both 0, $\gcd(a, b)$ is a linear combination of a and b.*

Proof. As above, we can concentrate on the case $a > b > 0$. The proof proceeds by strong induction on the value of a. In the base case, $a = 2$, $b = 1$, and $\gcd(a, b) = 1 = 0 \cdot a + 1 \cdot b$. Assume that the theorem holds for all pairs a, b with $0 < b < a \leq k$. Consider a pair a', b' with $0 < b' < a' = k + 1$. If $b' | a'$ then $\gcd(a', b') = b'$ and the theorem trivially holds. Otherwise use the division algorithm to express $a' = qb' + r$, where $0 < r < b'$. By the induction hypothesis, there exist coefficients u and v, such that $\gcd(b', r) = ub' + vr$. Lemma 4.3.1 shows that $\gcd(a', b') = \gcd(b', r)$, therefore $\gcd(a', b') = ub' + vr = ub' + v(a' - qb') = va' + (u - vq)b'$. This shows that $\gcd(a', b')$ is a linear combination of a' and b' and completes the proof by induction. □

Bezout's identity implies that the set of linear combinations of a and b is the same as the set of multiples of their greatest common divisor (!):

Corollary 4.4.2. *An integer z is a linear combination of a and b if and only if it is a multiple of $\gcd(a, b)$. In particular, $\gcd(a, b)$ is the least positive linear combination of a and b.*

Proof. By Theorem 4.2.1(c), since $\gcd(a, b)$ divides both a and b, it divides any linear combination z of a and b, and thus z is a multiple of $\gcd(a, b)$. On the other hand, we know by Bezout's identity that there are coefficients u and v, such that $\gcd(a, b) = ua + vb$, so if $z = c \cdot \gcd(a, b)$, then $z = c(ua + vb) = (cu)a + (cu)v$. □

Chapter 5

Prime Numbers

5.1 The fundamental theorem of arithmetic

Definition: An integer $p > 1$ is said to be *prime* if its only positive divisors are 1 and p itself. All other integers greater than 1 are called composite.

A composite number n can be written as a product $n = ab$ of two strictly smaller numbers $1 < a, b < n$. Note that, by convention, 1 is neither prime nor composite. Here are all primes below 100:

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97.$$

Given a prime p and another integer a, either a is a multiple of p or $\gcd(p, a) = 1$. Indeed, $\gcd(p, a)$ divides p, so it must be either 1 or p, and since $\gcd(p, a)$ also divides a then either $\gcd(p, a) = 1$ or a is a multiple of p. This can be used to prove a very important property of primes:

Theorem 5.1.1. *Let p be a prime.*

(a) Given two integers a and b, if $p|ab$ then either $p|a$ or $p|b$.

(b) Given k integers $a_1, a_2, \ldots, a_k$, if $p| \prod_{i=1}^{k} a_i$ then $p|a_i$ for some $1 \leq i \leq k$.

Proof.

(a) If $p|a$ we are done. Otherwise $\gcd(p, a) = 1$ and by Bezout's identity there exist linear coefficients u and v for which $1 = ua + vp$. Multiplying both sides by b we get $b = uab + vpb$. Since p divides ab, p divides the whole sum $uab + vpb$. Therefore $p|b$.

(b) The proof proceeds by induction. The case $k = 1$ is trivial and $k = 2$ is handled in part (a). So we assume that the claim holds for some $k > 1$ and prove that it also holds for $k + 1$. Given that $p| \prod_{i=1}^{k+1} a_i$, we put $b = \prod_{i=1}^{k} a_i$. Since $p|ba_{k+1}$, part (a) implies that either $p|a_{k+1}$ or $p|b$. In both cases the claim holds, in the latter case by the induction hypothesis. This completes the proof by induction.

□

Theorem 5.1.1 can be used to derive a fundamental theorem of number theory. It is so fundamental it has "fundamental" in its name.

Theorem 5.1.2 (Fundamental Theorem of Arithmetic)**.** *Every positive integer can be represented in a unique way as a product of primes,*

$$n = p_1 p_2 \cdots p_k \qquad (p_1 \leq p_2 \leq \ldots \leq p_k).$$

Proof. We first prove existence and then uniqueness. Actually, we already proved existence in one of the previous lectures as an illustration of strong induction, but give the prove here again for completeness. So, to prove that every integer can be represented as a product of primes we use strong induction. The base case $n = 1$ holds because the *empty product*, as we previously discussed, is defined to equal 1. The induction hypothesis assumes that for some $n > 1$, all positive integers $k < n$ can be represented as a product of primes. If n is prime, then it is trivially a product of primes. Otherwise it can be written as $n = ab$, for $1 < a, b < n$. By the induction hypothesis, both a and b are products of primes, so their product n is also a product of primes. This proves existence.

The proof that the above representation is unique proceeds by contradiction. Assume then that there exists some positive integer that can be represented as a product of primes in (at least) two ways. By the well-ordering principle, there is a smallest such integer n. It holds that $n = p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_l$, where $p_1 \leq p_2 \leq \ldots \leq p_k$, $q_1 \leq q_2 \leq \ldots \leq q_l$, and $p_i \neq q_i$ for some i. By Theorem 5.1.1(b), since $p_i | q_1 q_2 \cdots q_l$, there must exist some q_j for which $p_i | q_j$. Since q_j is prime and $p_i > 1$, this can only occur when $p_i = q_j$. Thus we can eliminate p_i and q_j from the equation $p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_l$ and get two distinct representations of the positive integer number n/p_i as a product of primes. This contradicts the assumption that n is the smallest positive integer with this property, and concludes the proof of uniqueness. □

5.2 The infinity of primes

Here is another fundamental result with a proof from Euclid's *Elements*:

Theorem 5.2.1. *There are infinitely many primes.*

Proof. Assume for the sake of contradiction that there is only a finite set of primes, $p_1, p_2, \ldots, p_n$. Consider the number

$$p = p_1 p_2 \ldots p_n + 1.$$

By Theorem 5.1.2, p has a prime divisor, which has to be p_i, for some $1 \leq i \leq n$. Since p_i divides both p and $p_1 p_2 \ldots p_n$, it also divides $p - p_1 p_2 \ldots p_n = 1$. However, this is impossible since $p_i > 1$. This contradiction proves the theorem. □

Let's get some more mileage out of Euclid's proof. The results below show that not only do the primes never stop, but the number of primes $p \leq x$ is at least a certain natural function of x, namely at least $\log \log x$. (Here the base of the logarithm is 2.)

Theorem 5.2.2. *The n-th prime p_n satisfies $p_n \leq 2^{2^{n-1}}$ for all $n \geq 1$.*

Proof. We proceed using strong induction. For the base case, the first prime is $2 = 2^{2^0}$. Assume that the claim holds for all primes p_1 through p_k. Consider $p = p_1 p_2 \dots p_k + 1$. As in the above proof, p has a prime factor that is not one of the first k primes. This prime factor is thus at least as large as p_{k+1}, which implies

$$\begin{aligned} p_{k+1} \leq p = p_1 p_2 \dots p_k + 1 &\leq 2^{2^0} 2^{2^1} \cdots 2^{2^{k-1}} + 1 \\ &= 2^{1+2+4+\dots+2^{k-1}} + 1 \\ &= 2^{2^k - 1} + 1 \\ &= \frac{1}{2} 2^{2^k} + 1 \\ &\leq 2^{2^k}. \end{aligned}$$

This is precisely the induction step we needed, and concludes the proof by strong induction. □

Denote by $\pi(x)$ the number of primes $p \leq x$.

Corollary 5.2.3. *For $x \geq 2$, $\pi(x) \geq \lfloor \log \log x \rfloor + 1$.*

Proof. Plugging $n = \lfloor \log \log x \rfloor + 1$ into Theorem 5.2.2 implies that the n-th prime is at most x. Thus there are at least n primes below x. □

For general education, you should know that this is by far not the best possible estimate. A celebrated achievement in number theory is the Prime Number Theorem due to Hadamard and de la Vallée Poussin, which states that $x/\ln x$ (here we use the natural logarithm) is the "right" bound, in the sense that

$$\lim_{x \to \infty} \frac{\pi(x)}{x/\ln x} \to 1.$$

Chapter 6

Modular Arithmetic

6.1 Congruences

We usually associate arithmetic with the infinite set of integer numbers. However, *modular arithmetic* on finite sets is commonly used in our daily life. As an example, if it is now 1 am and we let 1000 hours pass, what time will it be? We can use the division algorithm to see that $1000 = 41 \times 24 + 16$ and conclude that adding 1000 hours is like adding 16 hours, since the clock returns to the same position every 24 hours. So after 1000 hours it will be 5 pm (17 hours after midnight).

There are many examples in which it is natural and useful to limit our number system to a finite range of integers, such as 0 through $n-1$, for some n. This number system is denoted by $\mathbb{Z}_n$. Days of the week, hours of the day, minutes in an hour are all familiar examples of finite number systems, as are numbers in microprocessor registers, commonly limited to 32 binary digits.

Modular arithmetic allows us to add, subtract, multiply, and sometimes divide numbers while staying within the finite set $\mathbb{Z}_n$. The number n is called the *modulus*. A central notion in modular arithmetic is *congruence*. We say that two integers are congruent modulo n if they leave the same remainder when divided by n. Here is the formal definition:

Definition: Two integers $a, b \in \mathbb{Z}$ are said to be *congruent modulo* n, written as $a \equiv_n b$ or $a \equiv b \pmod{n}$, if and only if they leave the same remainder when divided by n, that is, $a \text{ rem } n = b \text{ rem } n$.

This definition captures our intuition that the day of the week will be the same whether we let 10, 17, or 80 days pass. There is an equivalent definition of congruence that is often useful in proofs:

Lemma 6.1.1. *$a \equiv_n b$ if and only if $n|(a-b)$.*

Proof. If $a \equiv_n b$ then $a \text{ rem } n = b \text{ rem } n$. Put $r = a \text{ rem } n = b \text{ rem } n$. Then there exist two integers q_1 and q_2, such that $a = q_1 n + r$ and $b = q_2 n + r$. Subtracting the second equation from the first, we get $a - b = (q_1 - q_2)n$ and $n|(a-b)$.

On the other hand, if $n|(a-b)$ then there exists an integer d, such that $a - b = nd$. By the division algorithm, there exist integers $q_1, q_2 \in \mathbb{Z}$, and $0 \leq r_1, r_2 < n$, such

that $a = q_1 n + r_1$ and $b = q_2 n + r_2$. Thus $(q_1 - q_2)n + (r_1 - r_2) = nd$, and $r_1 - r_2 = (q_2 - q_1 + d)n$. Thus $n|(r_1 - r_2)$. However, $|r_1 - r_2| < n$, so necessarily $r_1 - r_2 = 0$, which implies that $a \text{ rem } n = b \text{ rem } n$, and $a \equiv_n b$. □

You should use the definition to verify that for any $a, b, c \in \mathbb{Z}$,

- $a \equiv_n a$. (Reflexivity.)
- If $a \equiv_n b$ then $b \equiv_n a$. (Symmetry.)
- If $a \equiv_n b$ and $b \equiv_n c$ then $a \equiv_n c$. (Transitivity.)

The operations of addition, subtraction, and multiplication on $\mathbb{Z}_n$ are defined by first doing the corresponding operation in $\mathbb{Z}$ and then taking the remainder modulo n. That is, if we denote these respective operations by $+_n$, $-_n$, and $\cdot_n$, then

$$\begin{aligned} a +_n b &= (a + b) \text{ rem } n \\ a -_n b &= (a - b) \text{ rem } n \\ a \cdot_n b &= (ab) \text{ rem } n \end{aligned}$$

Exponentiation is defined through repeated multiplication.

Lemma 6.1.2. *Properties of congruence:*

(a) $(a \text{ rem } n) \text{ rem } n = a \text{ rem } n$

(b) $(a \text{ rem } n) \equiv_n a$

(c) $(ab) \text{ rem } n = (a \text{ rem } n)(b \text{ rem } n) \text{ rem } n$

(d) $(a \text{ rem } n)(b \text{ rem } n) \equiv_n ab$

(e) $\prod_{i=1}^{k} (a_i \text{ rem } n) \equiv_n \prod_{i=1}^{k} a_i$

(f) If $a_1 \equiv_n a_2$ *and* $b_1 \equiv_n b_2$ *then*

$$\begin{aligned} a_1 + b_1 &\equiv_n a_2 + b_2 \\ a_1 - b_1 &\equiv_n a_2 - b_2 \\ a_1 b_1 &\equiv_n a_2 b_2 \end{aligned}$$

Proof. (b) is just a restatement of (a). To prove these we need to show that $n|(a - (a \text{ rem } n))$. Put $r = a \text{ rem } n$. By the division algorithm, there exists $q \in \mathbb{Z}$, such that $a = qn + r$. Thus $a - r = qn$, which implies that $n|a - r$ and concludes the proof.

(d) is a restatement of (c), and (e) can be proved from (d) by induction. To prove (c) we need to show that $n|(ab - (a \text{ rem } n)(b \text{ rem } n))$. Use the division algorithm to represent $a = q_1 n + r_1$ and $b = q_2 n + r_2$. Then

$$ab - (a \text{ rem } n)(b \text{ rem } n) = (q_1 n + r_1)(q_2 n + r_2) - r_1 r_2 = (q_1 q_2 n + r_1 q_2 + q_1 r_2)n,$$

which implies the claim.

We now prove (f). We know that $n|(a_1 - a_2)$ and $n|(b_1 - b_2)$. That is, there exist integers q and s, such that $a_1 - a_2 = qn$ and $b_1 - b_2 = sn$. Adding these equations gives $(a_1 + b_1) - (a_2 + b_2) = (q + s)n$, which yields the first part of the claim. Subtracting similarly gives the second part. Writing $a_1 = a_2 + qn$ and $b_1 = b_2 + sn$ and multiplying the equations gives

$$\begin{aligned} a_1 b_1 &= a_2 b_2 + b_2 qn + a_2 sn + qsn^2 \\ a_1 b_1 - a_2 b_2 &= (b_2 q + a_2 s + qsn)n, \end{aligned}$$

which yields the third part. □

6.2 Modular division

You might have noticed that we defined addition, subtraction, and multiplication, but not division. This might not be surprising, since the division operation is not defined for the integers in general: There is no integer that corresponds to 5 divided by 4, for instance. (In other words, there is no $x \in \mathbb{Z}$, such that $4x = 5$.) This distinguishes $\mathbb{Z}$ from sets like $\mathbb{Q}$ or $\mathbb{R}$ that are *closed under division*.

Division in $\mathbb{Z}_n$ appears even more unruly. For example, in $\mathbb{Z}_6$, the equation $2x = 4$ is satisfied by both $x = 2$ and $x = 5$, while the equation $2x = 3$ has no solutions. So the notion of "b divided by a" can be undefined or even ambiguous in $\mathbb{Z}_n$. In particular, we cannot generally cancel a multiplier from both sides of a congruence, that is, if $ab \equiv_n ac$ we cannot reason that $b \equiv_n c$. To take the above illustration, $2 \cdot 2 \equiv_6 2 \cdot 5$, but $2 \not\equiv_6 5$.

Quite remarkably, however, the division operation is well-defined when n is a prime p. Thus $\mathbb{Z}_p$ is in a sense as well-behaved as the real numbers, despite being a finite set! After a small digression that explores what "well-behaved" actually means here, we will state an even more general result on modular division.

Digression (notions from abstract algebra): There is a way to precisely state what we mean by "well-behaved" above. Jumping the gun, I'll say that $\mathbb{Z}_p$ is a *field*, not just a *ring*. Now let me tell you what this means. The notion of a ring in algebra is meant to abstract our intuition concerning the essential properties of the integers. Given a set S equipped with two operations, $+$ (addition) and $\cdot$ (multiplication), we say that S is a ring if the following all hold for any $a, b, c \in S$:

- $a + b \in S$ and $a \cdot b \in S$.
- $a + (b + c) = (a + b) + c$ and $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
- $a + b = b + a$ and $a \cdot b = b \cdot a$.
- $a \cdot (b + c) = a \cdot b + a \cdot c$.
- There exists an *additive identity* element $0 \in S$ that satisfies $a + 0 = a$ and a *multiplicative identity* element $1 \in S$ that satisfies $a \cdot 1 = a$ for all $a \in S$.

- For every $a \in S$ there exists an *additive inverse* $-a \in S$ for which $a + (-a) = 0$.

All the number systems we have encountered so far are rings, including $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{Z}_n$. However, some of them possess additional structure that allows the division operation. Namely, a ring is said to be a *field* if, in addition to the above, the following holds

- For every $a \in S$, such that $a \neq 0$, there exists a *multiplicative inverse* $a^{-1} \in S$ for which $a \cdot a^{-1} = 1$.

The number systems $\mathbb{R}$ and $\mathbb{Q}$, as well as $\mathbb{Z}_p$ when p is prime, are fields. In fields the division operation is well-defined, and $b/a = b \cdot a^{-1}$, as can be verified by plugging $x = b \cdot a^{-1}$ into the equation $ax = b$. A field with a finite number of elements is called a *Galois field*, after the French mathematician Evariste Galois. (A feisty young man who died in a duel at the age of 20, *after* making significant enough contributions to mathematics to have a whole field (sic) named in his honor!) Anyway, now that we know what fields are, let's see why $\mathbb{Z}_p$ is one. In fact, we prove something more general:

Theorem 6.2.1. *If a and n are coprime then there exists exactly one $x \in \mathbb{Z}_n$ for which $ax \equiv_n b$, for any $b \in \mathbb{Z}$.*

Proof. We need to prove existence and uniqueness of x as described in the theorem. $ax \equiv_n b$ if and only if there exists $q \in \mathbb{Z}$, such that $ax - b = nq$, or $ax - nq = b$. Now, since $\gcd(a, n) = 1$, any integer, including b, is a linear combination of a and n. This proves existence.

To prove uniqueness, assume that for $x, y \in \mathbb{Z}_n$ it holds that $ax \equiv_n b$ and $ay \equiv_n b$. Thus $ax - ay \equiv_n 0$, or $n|a(x-y)$. As you proved in one of the homework assignments, since n and a are coprime, this implies that $n|(x-y)$, and therefore that $x - y \equiv_n 0$. Thus $x \equiv_n y$, which proves uniqueness. □

Corollary 6.2.2. *For a prime p and any $a, b \in \mathbb{Z}$, such that $a \not\equiv_p 0$, there exists exactly one $x \in \mathbb{Z}_p$ for which $ax \equiv_p b$.*

The fact that division is well-defined in $\mathbb{Z}_p$ when p is prime also means that cancelations become valid. Thus if $a \not\equiv_p 0$ and $ab \equiv_p ac$ we can safely conclude that $b \equiv_p c$.

We now know that b/a is well-defined in $\mathbb{Z}_p$, but how do we find it? That is, how do we find $x \in \mathbb{Z}_p$, for which $ax \equiv_p b$. This question is particularly important when p is large and it takes too long to simply enumerate all the elements of $\mathbb{Z}_p$. Fortunately, the following result, known as *Fermat's Little Theorem*, can help us:

Theorem 6.2.3. *For a prime p and any $a \not\equiv_p 0$,*

$$a^{p-1} \equiv_p 1.$$

Proof. Consider the set S, defined as $1 \cdot a, 2 \cdot a, \ldots, (p-1) \cdot a$. None of these $p - 1$ integers are congruent modulo p, since we have seen that if $ia \equiv_p ja$ then $i \equiv_p j$.

However, each element of S is congruent to some element of $\mathbb{Z}_p$. Since there are $p-1$ elements in S and $p-1$ nonzero elements in $\mathbb{Z}_p$, the elements of S must be congruent to each of $1, 2, \ldots, (p-1)$ in some order. Therefore,

$$1 \cdot 2 \cdot \cdots \cdot (p-1) \equiv_p 1a \cdot 2a \cdot \cdots \cdot (p-1)a,$$

or

$$1 \cdot 2 \cdot \cdots \cdot (p-1) \equiv_p 1 \cdot 2 \cdot \cdots \cdot (p-1) \cdot a^{p-1}.$$

We can cancel each of $1, 2, \ldots, (p-1)$ from both sides of the congruence, obtaining $a^{p-1} \equiv_p 1$. □

Fermat's Little Theorem allows us to quickly perform division in $\mathbb{Z}_p$. The element $x \in \mathbb{Z}_p$ for which $ax \equiv_p b$ is simply $(a^{p-2}b \text{ rem } p)$.