THE INTEGERS

1. Divisibility and Factorization

Without discussing foundational issues or even giving a precise definition, we take basic operational experience with the integers for granted. Specifically the set of integers is notated

$$\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\},$$

with no concern—at least for now—about the meaning of the ellipsis "…". However, the point to be emphasized is that we view the integers not only as a set but as a set equipped with two operations, addition and multiplication. That is, the integers are an *algebraic structure*. Really the structure is $(\mathbb{Z}, +, \cdot)$ rather than merely $\mathbb{Z}$, but once we are aware that the currency of algebra is structures rather than sets, such notation is pointlessly cumbersome.

The product of any two nonzero integers is again nonzero. Consequently if the product of a nonzero integer with a second integer is zero then the second integer is itself zero. This observation leads to the **cancellation law**:

$$\text{For all } a, b, c \in \mathbb{Z}, \text{ if } ab = ac \text{ and } a \neq 0 \text{ then } b = c.$$

Indeed, the given equality says that $a(b - c) = 0$, and $a \neq 0$, so $b - c = 0$.

The first substantive result about the integers is the division algorithm, going back to Euclid.

Theorem 1.1. *Let a and b be integers with $b \neq 0$. There exist unique integers q and r such that*

$$a = qb + r, \quad 0 \leq r < |b|.$$

Proof. A fundamental property of the integers is that any nonempty set of nonnegative integers contains a least element. The set

$$\{a - qb : q \in \mathbb{Z}\}$$

is readily seen to contain nonnegative elements (e.g., if $b > 0$ then $a - qb \geq 0$ for all integers $q \leq a/b$), so it contains a least nonnegative element $r = a - qb$, $r \geq 0$. Thus $r - |b| < 0$. The displayed conditions in the theorem follow. As for uniqueness, suppose that

$$qb + r = q'b + r', \quad r, r' \in \{0, \cdots, |b| - 1\}.$$

Then

$$r' - r = (q - q')b, \quad r' - r \in \{-|b| + 1, \cdots, |b| - 1\}.$$

The only multiple of b in $\{-|b| + 1, \cdots, |b| - 1\}$ is 0. Thus $r' = r$ and then $q = q'$ by the cancellation law. □

Let a and b be integers. If $a = qb$ for some integer q then b **divides** a and a is **divisible** by b. This condition is notated

$$b \mid a.$$

(For example, the only integer divisible by 0 is 0. However, since $0 = q0$ for *every* integer q, no meaningful rational value can be assigned to the "quotient" $0/0$.)

Proposition 1.2. *Let a and b be integers, not both zero. Consider the set of $\mathbb{Z}$-linear combinations of a and b,*

$$I(a,b) = \{ka + \ell b : k, \ell \in \mathbb{Z}\}.$$

Let

$$g = \textit{the least positive element of } I(a,b).$$

Then g is the **greatest common divisor** *of a and b, meaning that*

$$g \mid a, \qquad g \mid b, \qquad (d \mid a \textit{ and } d \mid b) \implies d \mid g.$$

The notation is

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

Proof. Easy calculations show that $I(a,b)$ is closed under $\mathbb{Z}$-linear combination,

$$\left\{\begin{array}{c} (ka+\ell b) + (k'a + \ell' b) = (k+k')a + (\ell + \ell')b \\ m(ka+\ell b) = (mk)a + (m\ell)b \end{array}\right\} \quad \text{for all } k, k', \ell, \ell', m \in \mathbb{Z}.$$

(Here the slogan is *Linear combinations of linear combinations are linear combinations.*) Write $a = qg + r$ where $0 \le r < g$. Since $r = a - qg$ lies in $I(a,b)$ and $0 \le r < g$, the definition of g as the least positive element of $I(a,b)$ shows that $r = 0$. Thus $g \mid a$, and similarly $g \mid b$.

For any integer d, even-easier calculations show that the set $I(d) = d\mathbb{Z}$ of $\mathbb{Z}$-linear combinations of d is closed under $\mathbb{Z}$-linear combination,

$$\left\{\begin{array}{c} kd + k'd = (k+k')d \\ m(kd) = (mk)d \end{array}\right\} \quad \text{for all } k, k', m \in \mathbb{Z}.$$

But $I(d)$ is the set of integers that d divides. So if $d \mid a$ and $d \mid b$ then $d \mid g$ since g is a $\mathbb{Z}$-linear combination of a and b. □

The letter I in the notations $I(a,b)$ and $I(d)$ stands for the mathematical term *ideal*. In mathematical parlance *ideal* is a noun. Later in the course we will learn explicitly about ideals, but already here it is worth noting that the ideals $I(a,b)$ and $I(d)$ are algebraic structures having the property of closure under linear combinations whereas the sets $\{a,b\}$ and $\{d\}$ have no such property. The closure property of the algebraic structures $I(a,b)$ and $I(d)$ is what makes the flow of ideas in the proof so smooth.

As an example of finding a greatest common divisor, abbreviate the notations $I(a,b)$ and $I(d)$ to (a,b) and (d), and compute

$$\begin{aligned} (826, 1890) &= (826, 1890 - 2 \cdot 826) \\ &= (238, 826) = (238, 826 - 3 \cdot 238) \\ &= (112, 238) = (112, 238 - 2 \cdot 112) \\ &= (14, 112) = (14, 112 - 8 \cdot 14) \\ &= (0, 14) = (14). \end{aligned}$$

Thus $\gcd(826, 1890) = 14$. The process just demonstrated is the *Euclidean algorithm*. And furthermore, we can backtrack to express the gcd as a linear combination of the two given numbers,

$$\begin{aligned} 14 &= 238 - 2 \cdot 112 \\ &= 238 - 2 \cdot (826 - 3 \cdot 238) \\ &= 7 \cdot 238 - 2 \cdot 826 \\ &= 7 \cdot (1890 - 2 \cdot 826) - 2 \cdot 826 \\ &= -16 \cdot 826 + 7 \cdot 1890. \end{aligned}$$

This process shows that we know how to solve any equation of the form

$$ax + by = c,$$

where $a, b, c \in \mathbb{Z}$ are the given coefficients and we seek integer solutions (x, y). Solutions exist if and only if $\gcd(a, b) \mid c$, in which case we can find one particular solution via the Euclidean algorithm, as above. All other solutions differ from the particular solution by solutions to the homogenized equation $ax + by = 0$, which is easy to solve: after dividing a and b by their gcd we get $a'x + b'y = 0$ where $\gcd(a', b') = 1$, and so the solutions are $(x, y) = n(b', -a')$ for all $n \in \mathbb{Z}$.

Definition 1.3. *Let p be a positive integer greater than 1. Then*

- *p is* **irreducible** *if*

  $$\textit{the only positive divisors of } p \textit{ are } 1 \textit{ and } p.$$

- *p is* **prime** *if*

  $$\textit{for all } a, b \in \mathbb{Z}, \quad p \mid ab \implies p \mid a \textit{ or } p \mid b.$$

Note that *irreducible* means to us what *prime* means to most people, while *prime* means to us something else, perhaps unfamiliar. However, we next show that in the context of the integers, the two words mean the same thing after all.

Proposition 1.4. *Let p be a positive integer greater than 1. Then p is irreducible if and only if p is prime.*

Proof. Let p be irreducible, and suppose that a and b are integers such that $p \mid ab$. If $p \mid a$ then we are done. If $p \nmid a$ then p and a share no positive divisor except 1, and their greatest common divisor 1 is a linear combination of them,

$$1 = kp + \ell a \quad \text{for some } k, \ell \in \mathbb{Z}.$$

Consequently

$$b = kpb + \ell ab \quad \text{for some } k, \ell \in \mathbb{Z}.$$

Because $p \mid ab$ by hypothesis, the right side is divisible by p, and hence so is the left side. That is, if $p \mid ab$ and $p \nmid a$ then $p \mid b$ as desired.

Let p be prime and suppose that d is a positive divisor of p. Thus $kd = p$ for some positive integer k, and so $p \mid k$ or $p \mid d$.

- If $k = ep$ then the equality $kd = p$ becomes $edp = p$, so that $ed = 1$, forcing $d = 1$.
- If $d = ep$ then the equality $kd = p$ becomes $kep = p$, so that $ke = 1$, forcing $k = 1$ and thus $d = p$.

$\square$

The proof that irreducible implies prime requires the division algorithm, but the proof that prime implies irreducible does not. Later in the course we will revisit these issues in more generality. In many environments that are otherwise similar to the integers, irreducible does not imply prime.

The equivalence of irreducibility and primality in $\mathbb{Z}$ is necessary to prove that every nonzero integer factors uniquely as a sign-term times a product of irreducibles.

Theorem 1.5 (Unique Factorization of Integers)**.** *Every nonzero integer has a unique factorization*

$$n = \pm p_1^{e_1} \cdots p_g^{e_g}, \quad g \geq 0,\ p_1 < \cdots < p_g,$$

where the p_i are irreducible.

Proof. We may assume that n is positive.

(Existence.) The integer $n = 1$ takes the desired form. For $n > 1$, if n is irreducible then it takes the desired form. Otherwise n factors as $n = n_1 n_2$ where the positive integers n_1 and n_2 are smaller than n, so each of them is a product of finitely many irreducibles by induction, and hence so is n. If $n < 0$ then either $n = -1$ or $-n$ is a product of finitely many irreducibles.

(Uniqueness.) It suffices to consider positive integers. The only case to worry about is two nontrivial factorizations,

$$n = \prod_{i=1}^{g} p_i^{e_i} = \prod_{j=1}^{h} q_j^{f_j}, \quad g, h \geq 1.$$

Since p_1 divides the first product it divides the second one, and hence (p_1 being prime) it divides one of the q_j, and hence (q_j being irreducible) it equals q_j. And q_j must be the smallest prime on the right side, for otherwise q_1, which by the argument just given must equal some p_i, can not do so, being smaller than $q_j = p_1$. Thus $p_1 = q_1$. To see that $e_1 = f_1$, divide the equality through by $\min\{e_1, f_1\}$ to get an equality in which p_1 doesn't divide one side, hence doesn't divide the other side either, giving the result. Now we have an equality

$$n/p_1^{e_1} = \prod_{i=2}^{g} p_i^{e_i} = \prod_{j=2}^{h} q_j^{f_j},$$

and we are done by induction on n. □

Some of the issues here may seem needlessly complicated, but they are forced on us by familiar elementary contexts. For example:

- Math 112 exercises have to dance around the question

  For what positive integers n is $\mathbb{Z}/n\mathbb{Z}$ a field?

  Everybody knows morally that the answer is, *For prime n.* However, the problem is that for any $n \in \mathbb{Z}^+$, short, natural arguments show that

  $$n \text{ is prime} \iff \mathbb{Z}/n\mathbb{Z} \text{ is a field} \implies n \text{ is irreducible}.$$

  But to show that $\mathbb{Z}/n\mathbb{Z}$ is a field *only* if n is irreducible requires the fact that irreducibles are prime in $\mathbb{Z}$, which in turn relies on the division algorithm and the expression of the gcd as a linear combination.
- Similarly, any argument that there is no square root of 2 in $\mathbb{Q}$ tacitly makes use of unique factorization. Ultimately the argument boils down to

> *Let $r \in \mathbb{Q}$ satisfy $r^2 = 2$. Then $r = 2^e r'$ where $e \in \mathbb{Z}$ and $r' \in \mathbb{Q}$ has no 2's in its numerator or its denominator. Thus $2 = r^2 = 2^{2e}(r')^2$. But this is impossible because there are no 2's in $(r')^2$ and so the left side has one power of 2 while the right side has an even number of 2's.*

The problem with the argument is that without unique factorization, a number with one power of 2 conceivably could equal a number with an even number of 2's. Any side-argument that no integer can be simultaneously odd and even must essentially give the argument that the division algorithm exists, specialized to $b = 2$.

2. The Integers modulo n

Let n be a positive integer. We want to define *clock algebra* modulo n, meaning addition and multiplication that wrap around at n, returning to 0. Thus it is initially tempting to work with the set of nonnegative remainders modulo n,

$$\{0, 1, \cdots, n-1\},$$

and to define a new addition $\oplus$ on this set by the rule

$$a \oplus b = r \quad \text{where } a + b = qn + r \text{ and } 0 \le r < n,$$

and similarly for multiplication. Although this idea can be made to work, it is the wrong idea. For one thing, carrying out the division algorithm for each addition or multiplication is constraining—we might prefer to throw away all multiples of n at the end of a long calculation instead. For another thing, if n is odd then it might be more natural to work symmetrically about 0 by using the set

$$\{-(n-1)/2, \cdots, -1, 0, 1, \cdots, (n-1)/2\}.$$

Now the wraparound occurs at $(n+1)/2$, returning to $-(n-1)/2$, but clearly the situation is essentially the same as wrapping around from n to 0.

The right idea is to view any two integers that differ by a multiple of n as equivalent, since the algebra of the integers gives rise to sensible algebra at the level of equivalence. The idea presents two psychological obstacles:

- The basic elements to be worked with are now *equivalence classes* such as

$$\begin{aligned} \{0, \pm n, \pm 2n, \cdots\} &= n\mathbb{Z}, \\ \{1, 1 \pm n, 1 \pm 2n, \cdots\} &= 1 + n\mathbb{Z}, \\ \{2, 2 \pm n, 2 \pm 2n, \cdots\} &= 2 + n\mathbb{Z}. \end{aligned}$$

  That is, each set in the display is to be treated as *one unit*, not as an infinitude.
- The question is not *how* to work with the equivalence classes, since of course the only feasible definitions are

$$\begin{aligned} (a + n\mathbb{Z}) \oplus (b + n\mathbb{Z}) &= a + b + n\mathbb{Z}, \\ (a + n\mathbb{Z}) \otimes (b + n\mathbb{Z}) &= ab + n\mathbb{Z}. \end{aligned}$$

  The question, which can take some thought to wrap one's mind around, is whether the definitions are *meaningful*. The problem is that while an integer a determines an equivalence class $a + n\mathbb{Z}$, an equivalence class arises from infinitely many choices of a; but the right sides in the previous display made specific choices of elements a and b in the two equivalence classes in order to define their sum and product.

(As an example of how a process's dependence on choices can render it ill-defined, suppose that we have a function

$$f : \{\text{Math 332 students}\} \times \{\text{Math 332 students}\} \longrightarrow \{\text{Math 332 students}\},$$

i.e., given an ordered pair of students, the mechanism specifies a student. Group students into equivalence classes by hair-color, and try to define a corresponding function

$$F : \{\text{Math 332 hair-colors}\} \times \{\text{Math 332 hair-colors}\} \longrightarrow \{\text{Math 332 hair-colors}\}$$

as follows:

- Given colors (c_1, c_2), pick equivalence class elements, a student s_1 having hair-color c_1 and a student s_2 having hair-color c_2.
- Let $s_3 = f(s_1, s_2)$, a student.
- Let $F(c_1, c_2)$ be the equivalence class (hair-color) of s_3.

Obviously $F(c_1, c_2)$ depends not only on c_1 and c_2 but also on the choices of s_1 and s_2. It is not a well-defined function of its two inputs.)

To show that the definitions do make sense, consider two different descriptions of each of two equivalence classes,

$$\begin{aligned} a + n\mathbb{Z} &= a' + n\mathbb{Z}, \\ b + n\mathbb{Z} &= b' + n\mathbb{Z}. \end{aligned}$$

This means that

$$\begin{aligned} a' - a &\in n\mathbb{Z}, \\ b' - b &\in n\mathbb{Z}, \end{aligned}$$

so that (since the ideal $n\mathbb{Z}$ is closed under linear combinations)

$$\begin{aligned} (a' + b') - (a + b) &= (a' - a) + (b' - b) \in n\mathbb{Z}, \\ a'b' - ab &= a'(b' - b) + (a' - a)b \in n\mathbb{Z}, \end{aligned}$$

and thus

$$\begin{aligned} a + b + n\mathbb{Z} &= a' + b' + n\mathbb{Z}, \\ ab + n\mathbb{Z} &= a'b' + n\mathbb{Z}. \end{aligned}$$

In other words, the sum and the product of the two equivalence classes is independent of the classes's descriptions,

$$\left\{ \begin{aligned} a + n\mathbb{Z} = a' + n\mathbb{Z} \\ b + n\mathbb{Z} = b' + n\mathbb{Z} \end{aligned} \right\} \implies \left\{ \begin{aligned} (a + n\mathbb{Z}) \oplus (b + n\mathbb{Z}) = (a' + n\mathbb{Z}) \oplus (b' + n\mathbb{Z}) \\ (a + n\mathbb{Z}) \otimes (b + n\mathbb{Z}) = (a' + n\mathbb{Z}) \otimes (b' + n\mathbb{Z}) \end{aligned} \right\}.$$

Although verifying that the definitions are sensible is tedious, the payoff is that their naturality guarantees that their algebra behaves well. For instance, the addition of equivalence classes inherits associativity from the addition of integers,

$$\begin{aligned}&((a+n\mathbb{Z})\oplus(b+n\mathbb{Z}))\oplus(c+n\mathbb{Z})\\&\qquad=(a+b+n\mathbb{Z})\oplus(c+n\mathbb{Z})\\&\qquad=(a+b)+c+n\mathbb{Z}\\&\qquad=a+(b+c)+n\mathbb{Z}\\&\qquad=(a+n\mathbb{Z})\oplus(b+c+n\mathbb{Z})\\&\qquad=(a+n\mathbb{Z})\oplus((b+n\mathbb{Z})\oplus(c+n\mathbb{Z})).\end{aligned}$$

And similarly for the distributive law,

$$\begin{aligned}&(a+n\mathbb{Z})\otimes((b+n\mathbb{Z})\oplus(c+n\mathbb{Z}))\\&\qquad=(a+n\mathbb{Z})\otimes(b+c+n\mathbb{Z})\\&\qquad=a(b+c)+n\mathbb{Z}\\&\qquad=ab+ac+n\mathbb{Z}\\&\qquad=(ab+n\mathbb{Z})\oplus(cc+n\mathbb{Z})\\&\qquad=((a+n\mathbb{Z})\otimes(b+n\mathbb{Z}))\oplus((a+n\mathbb{Z})\otimes(c+n\mathbb{Z})).\end{aligned}$$

Next we lighten the notation to hide the process of dragging around copies of $n\mathbb{Z}$ that are ultimately immaterial:

Let the symbol-string $a = b \bmod n$ *mean that* $n \mid b - a$.

It is straightforward to verify that equality modulo n is an **equivalence relation**,

(1) For all $a \in \mathbb{Z}$, $a = a \bmod n$.
(2) For all $a, b \in \mathbb{Z}$, if $a = b \bmod n$ then $b = a \bmod n$.
(3) For all $a, b, c \in \mathbb{Z}$, if $a = b \bmod n$ and $b = c \bmod n$ then $a = c \bmod n$,

and that the equivalence classes are exactly the sets that we have been manipulating tortuously. In the new notation, our verification that the operations modulo n are sensible showed that that for all integers a, a', b, b',

$$\left\{\begin{matrix} a = a' \bmod n \\ b = b' \bmod n \end{matrix}\right\} \implies \left\{\begin{matrix} a + b = a' + b' \bmod n \\ ab = a'b' \bmod n \end{matrix}\right\}.$$

That is, we may freely add and multiply modulo n with no regard to whether the inputs or the outputs are reduced back into $\{0, \cdots, n-1\}$ along the way. The results will always be equivalent, and we may reduce the answer at the end if we so choose.

Definition 2.1. *Let n be a positive integer. The set of equivalence classes of integers modulo n, along with the algebra* (*addition and multiplication, associativity and commutativity of both operations, additive and multiplicative identities, and additive inverse*) *that they inherit from the integers is called the* ring of integers modulo n *and denoted* $\mathbb{Z}/n\mathbb{Z}$.

(Again, $\mathbb{Z}/n\mathbb{Z}$ tacitly denotes $(\mathbb{Z}/n\mathbb{Z}, +, \cdot)$ or even $(\mathbb{Z}/n\mathbb{Z}, \oplus, \otimes)$.)

Thus *now* we may freely think of the elements of $\mathbb{Z}/n\mathbb{Z}$ as $\{0, 1, \cdots, n-1\}$ or as $\{-(n-1)/2, \cdots, (n-1)/2\}$, or as any other set of equivalence class representatives.

For example, a famous and difficult argument by Gauss involving an odd prime p uses nonzero representatives modulo p spaced four apart and straddling 0,

$$\{\pm 2, \pm 6, \pm 10, \cdots, \pm(2p-4)\}.$$

And if during the course of doing algebra we move outside our chosen set of representatives, we can safely continue to calculate and translate back into the set only when the calculation is complete.

Proposition 2.2. *Let n be a positive integer. An integer a is multiplicatively invertible modulo n if and only if* $\gcd(a,n)=1$.

Proof. To say that a is multiplicatively invertible modulo n is to say that

$$ab = 1 \bmod n \quad \text{for some integer } b,$$

which is to say that

$$n \mid ab-1 \quad \text{for some integer } b,$$

which is to say that

$$ab-1 = kn \quad \text{for some integers } b, k,$$

which is to say that

$$ab+kn = 1 \quad \text{for some integers } b, k.$$

The last condition is $\gcd(a,n)=1$. □

A bit strangely, 0 is multiplicatively invertible modulo 1. The tiny point here is that in the modulo 1 world *all* integers are equal, making $0=1$.

3. Fermat's Little Theorem and Euler's Generalization

Definition 3.1. *Euler's* **totient** *function,*

$$\varphi : \mathbb{Z}_{>0} \longrightarrow \mathbb{Z}_{>0},$$

is

$$\varphi(n) = \textit{the number of multiplicatively invertible elements in } \mathbb{Z}/n\mathbb{Z}.$$

Thus $\varphi(1)=1$ (as explained a moment ago at the end of the previous section) and $\varphi(p)=p-1$ if p is prime. Also, it is not hard to count Euler's totient function of a prime power by eliminating all multiples of the prime,

$$\varphi(p^e) = p^e - p^{e-1} = p^e(1-1/p), \quad p \text{ prime}, \; e \in \mathbb{Z}_{>0}.$$

Soon we will show that also

$$\varphi(mn) = \varphi(m)\varphi(n) \quad \text{if } \gcd(m,n)=1,$$

giving a complete formula for φ,

$$\boxed{\varphi(n) = n\prod_{p\mid n}(1-1/p).}$$

Let n be a positive integer. Consider some fixed invertible element a in $\mathbb{Z}/n\mathbb{Z}$. Let

$$\{x_1, x_2, \cdots, x_{\varphi(n)}\}$$

be the invertible elements of $\mathbb{Z}/n\mathbb{Z}$. Then also

$$\{ax_1, ax_2, \cdots, ax_{\varphi(n)}\}$$

is the same set. Indeed, each ax_i is invertible modulo n, the inverse of the product being the product of the inverses, and if $ax_i = ax_j \bmod n$ then multiplying through by the inverse of a modulo n gives $x_i = x_j \bmod n$. Thus multiplying the elements of the second set together produces the same result as multiplying the elements of the first set together,

$$a^{\varphi(n)}x = x \bmod n, \quad \text{where } x = x_1 x_2 \cdots x_{\varphi(n)}.$$

The element x is invertible modulo n, and so multiplying the equality through by its inverse gives

$$\boxed{a^{\varphi(n)} = 1 \bmod n \quad \text{if } \gcd(a,n) = 1.}$$

In particular,

$$\boxed{a^{p-1} = 1 \bmod p \quad \text{if } p \nmid a.}$$

The last equality is *Fermat's Little Theorem*, and the equality before it is *Euler's Generalization*.

Note that already in this first unit of the course, we are thinking in sufficiently structural terms that the generalization is more natural to prove than the original result. Fermat's Little Theorem can also be proved by induction: Certainly $1^{p-1} = 1 \bmod p$, and if $a^{p-1} = 1 \bmod p$ then $a^p = a \bmod p$ and so

$$\begin{aligned}(a+1)^p &= a^p + \binom{p}{1}a^{p-1} + \cdots + \binom{p}{p-1}a + 1 \\ &= a^p + 1 \bmod p \quad \text{since the binomial coefficients are } 0 \bmod p \\ &= a + 1 \bmod p \quad \text{since } a^p = a \bmod p.\end{aligned}$$

Thus, so long as $a+1$ hasn't reached $0 \bmod p$, cancellation gives $(a+1)^{p-1} = 1 \bmod p$ and the induction is complete. This argument feels more cluttered by auxiliary details than the previous one, but it points to a far-reaching ideas in its own right: if $p = 0$ then not only is the pth power of a product inevitably the product of the pth powers, $(ab)^p = a^p b^p$, but also the pth power of a *sum* is the sum of the pth powers, $(a+b)^p = a^p + b^p$. The fact that raising to the pth power preserves both algebra operations in any environment where $p = 0$ (not only in $\mathbb{Z}/p\mathbb{Z}$) leads to a great deal of rich mathematics.

Fermat's Little Theorem and Euler's Generalization would be facilitated by a fast raise-to-power method modulo n, i.e., a *fast modular exponentiation* algorithm. Such an algorithm is as follows. Let positive integers a, e, and n be given. The task is to compute the reduced power $a^e \,\%\, n$ quickly. (Here $x \,\%\, n$ denotes the element of $\{0, \cdots, n-1\}$ that equals x modulo n. Thus $x \,\%\, n$ is not an element of $\mathbb{Z}/n\mathbb{Z}$ since such elements are equivalence classes rather than class representatives.) We emphatically do not want to carry out $e - 1$ multiplications.

```
• (Initialize) Set $(x, y, f) = (1, a, e)$.
• (Loop) While $f > 0$, do as follows:
    – If $f\%2 = 0$ then replace $(x, y, f)$ by $(x, y^2 \,\%\, n, f/2)$,
    – otherwise replace $(x, y, f)$ by $(xy \,\%\, n, y, f - 1)$.
• (Terminate) Return $x$.
```

To see that the algorithm works by seeing how it works, represent the exponent e in binary, say

$$e = 2^g + 2^h + 2^k, \quad 0 \le g < h < k.$$

The algorithm successively computes

$$\begin{aligned}
&(1,\, a,\, 2^g + 2^h + 2^k)\\
&(1,\, a^{2^g},\, 1 + 2^{h-g} + 2^{k-g})\\
&(a^{2^g},\, a^{2^g},\, 2^{h-g} + 2^{k-g})\\
&(a^{2^g},\, a^{2^h},\, 1 + 2^{k-h})\\
&(a^{2^g+2^h},\, a^{2^h},\, 2^{k-h})\\
&(a^{2^g+2^h},\, a^{2^k},\, 1)\\
&(a^{2^g+2^h+2^k},\, a^{2^k},\, 0),
\end{aligned}$$

and then it returns the first entry, which is indeed a^e. The algorithm is strikingly efficient both in speed and in space. Especially, the operations on f (halving it when it is even, decrementing it when it is odd) are very simple in binary.

Fast modular exponentiation is not only for computers. For example, to compute $2^{37} \,\%\, 149$, proceed as follows,

$$\begin{aligned}
&(1,2;37) \to (2,2;36) \to (2,4;18) \to (2,16;9) \to (32,16;8)\\
&\to (32,-42;4) \to (32,-24;2) \to (32,-20;1) \to (105,-20;0).
\end{aligned}$$

And so the answer is 105.

As an example of using Fermat's Little Theorem and fast modular exponentiation, suppose that p is prime and $p = 3 \bmod 4$. Suppose that $a \neq 0 \bmod p$ is a square modulo p, that is, $a = b^2 \bmod p$ for some b, but we don't know what b is. Note that $p+1$ is divisible by 4, and compute, working modulo p and using Fermat's Little Theorem for the second-to-last equality, that

$$\left(a^{(p+1)/4}\right)^2 = \left(b^{(p+1)/2}\right)^2 = b^{p+1} = b^{p-1}b^2 = b^2 = a.$$

Thus $a^{(p+1)/4}$, which we can find quickly by fast modular exponentiation, is a square root of a modulo p. There is still the question of whether a given a has a square root modulo p at all. Answering that question quickly is a matter of the famous *Quadratic Reciprocity* theorem.

4. The Sun-Ze Theorem

Let m and n be positive integers whose greatest common divisor is 1. (Such a pair of integers is called *relatively prime* or *coprime*.) Associate to m and n the two algebraic structures

$$\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z} \quad \text{and} \quad \mathbb{Z}/mn\mathbb{Z}.$$

For the first structure, the algebraic operations are naturally defined componentwise (now writing "(m)" rather than "$\bmod m$" and likewise for n),

$$\begin{aligned}
(a\,(m), b\,(n)) + (a'\,(m), b'\,(n)) &= (a + a'\,(m), b + b'\,(n)),\\
(a\,(m), b\,(n))(a'\,(m), b'\,(n)) &= (aa'\,(m), bb'\,(n)).
\end{aligned}$$

Theorem 4.1 (Sun-Ze)**.** *Let m and n be coprime positive integers. Thus there exist integers k and ℓ such that*

$$km + \ell n = 1.$$

The maps

$$f : \mathbb{Z}/mn\mathbb{Z} \longrightarrow \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}, \quad a\,(mn) \longmapsto (a\,(m), a\,(n))$$

and

$$g : \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n \longrightarrow \mathbb{Z}/mn\mathbb{Z} \quad (a\,(m), b\,(n)) \longmapsto \ell na + kmb\,(mn),$$

are mutually inverse algebraic isomorphisms. That is, they are mutually inverse set-bijections, and they preserve the algebraic operations of the sets. (*The precise meaning of* preserve the operations *will be explained as part of the proof.*)

The Sun-Ze Theorem is often called the *Chinese Remainder Theorem*. A brief notation for its contents is

$$\mathbb{Z}/mn\mathbb{Z} \approx \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}.$$

Since the multiplicatively invertible elements of $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ are the pairs where each element is multiplicatively invertible, the Sun-Ze Theorem has in consequence a formula that was asserted earlier,

$$\varphi(mn) = \varphi(m)\varphi(n) \quad \text{if } \gcd(m, n) = 1,$$

Proof. The map

$$f : \mathbb{Z}/mn\mathbb{Z} \longrightarrow \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}, \quad a\,(mn) \longmapsto (a\,(m), a\,(n))$$

is meaningful: although we may translate a by any multiple of mn without affecting $a\,(mn)$, doing so has no affect on $a\,(m)$ or $a\,(n)$ either. To see that f of a sum of values in $\mathbb{Z}/mn\mathbb{Z}$ is the sum of the corresponding f-values in $\mathbb{Z}/m\mathbb{Z}\times\mathbb{Z}/n\mathbb{Z}$, compute

$$\begin{aligned} f(a\,(mn) + b\,(mn)) &= f(a + b\,(mn)) \\ &= (a + b\,(m), a + b\,(n)) \\ &= (a\,(m), a\,(n)) + (b\,(m), b\,(n)) \\ &= f(a\,(mn)) + f(b\,(mn)). \end{aligned}$$

And similarly for the product, i.e., $f(a\,(mn)\cdot b\,(mn)) = f(a\,(mn))\cdot f(b\,(mn))$, where the first product is set in $\mathbb{Z}/mn\mathbb{Z}$ and the second in $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$.

Recall the integers k, ℓ such that $km + \ell n = 1$. The map

$$g : \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n \longrightarrow \mathbb{Z}/mn\mathbb{Z} \quad (a\,(m), b\,(n)) \longmapsto \ell na + kmb\,(mn),$$

is also meaningful: translating a by any multiple of m translates $\ell na + kmb$ by a multiple of mn, and similarly for translating b by any multiple of n. To see that g preserves algebra just as f does, note first that because $km + \ell n = 1$, it follows that

$$k^2m^2 = km(1 - \ell n) = km \bmod mn,$$

and similarly $\ell^2 n^2 = \ell n \bmod mn$. Now compute,

$$\begin{aligned} g((a\,(m), b\,(n))(a'\,(m), b'\,(n))) &= g(aa'\,(m), bb'\,(n)) \\ &= \ell naa' + kmbb'\,(mn) \\ &= \ell^2n^2aa' + k^2m^2bb'\,(mn) \\ &= \big(\ell na + kmb\,(mn)\big)\big(\ell na' + kmb'\,(mn)\big) \\ &= g(a\,(m), b\,(n))\, g(a'\,(m), b'\,(n)). \end{aligned}$$

And similarly for the sum.

One composition of f and g is the identity on $\mathbb{Z}/mn\mathbb{Z}$,

$$g(f(a\,(mn))) = g(a\,(m), a\,(n))) = (\ell n + km) \cdot a\,(mn) = a\,(mn).$$

And since $km = 1 \bmod n$ and $\ell n = 1 \bmod m$, the other composition is the identity on $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$,

$$\begin{aligned} f(g(a\,(m), b\,(n))) &= f(\ell na + kmb\,(mn)) \\ &= (\ell na + kmb\,(m), \ell na + kmb\,(n)) \\ &= (a\,(m), b\,(n)). \end{aligned}$$

In sum, f and g preserve algebra and they are mutual inverses. Thus $\mathbb{Z}/mn\mathbb{Z}$ and $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ biject to one another as algebraic structures. □