

NUMBER SYSTEMS

Number theory is the study of the integers. We denote the set of integers by \mathbb{Z} :

$$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}.$$

The integers have two operations defined on them, addition and multiplication, which are commutative ($a + b = b + a$, $ab = ba$) and associative ($a + (b + c) = (a + b) + c$, $a(bc) = (ab)c$) and which interact via the distributive law ($a(b + c) = ab + ac$). These operations have neutral elements 0 and 1 respectively (if a is an integer, $a + 0 = a$ and $a \cdot 1 = a$). Notice also that each integer can be negated ($a + (-a) = 0$). In modern algebra language, a set having the aforementioned properties is called a *commutative ring*. The two operations in \mathbb{Z} are not “created equal”, however; while every integer can be negated (for example, $3 + (-3) = 0$), not every integer can be inverted (for example, there is no integer a such that $3a = 1$). Indeed, the only integers whose reciprocals are also integers are 1 and -1 . In general, an element a of a commutative ring is called a *unit* if there is an element b of the ring such that $ab = 1$.

The *rational numbers*, denoted by \mathbb{Q} , are all the ratios of integers:

$$\mathbb{Q} = \left\{ \frac{a}{b} \mid a, b \in \mathbb{Z} \text{ and } b \neq 0 \right\}$$

(of course, we consider $\frac{4}{6}$ and $\frac{2}{3}$ to be the same element of \mathbb{Q}). Notice that, like \mathbb{Z} , \mathbb{Q} is a commutative ring, but that any nonzero element of \mathbb{Q} can be inverted (if $\frac{a}{b} \in \mathbb{Q}$ and $a \neq 0$, then certainly $\frac{b}{a} \in \mathbb{Q}$ also). Commutative rings having this additional property are called *fields*.

Another field you are familiar with is the real numbers, which we will denote by \mathbb{R} . Notice that $\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$. We know how \mathbb{Z} sits in \mathbb{R} (imagine a number line with the integers marked off). You may have thought less about how \mathbb{Q} sits in \mathbb{R} .

Proposition 1. *Between any two real numbers, there is a rational number.*

Proof. Let $a, b \in \mathbb{R}$, $a < b$. We can find a positive integer n such that $\frac{1}{n} < b - a$. Then at least one of the rational numbers

$$\left\{ \dots, \frac{-2}{n}, \frac{-1}{n}, 0, \frac{1}{n}, \frac{2}{n}, \dots \right\}$$

lies between a and b . □

Because of Proposition 1, we say that \mathbb{Q} is *dense* in \mathbb{R} . However, not all real numbers are rational; a real number which is not rational is called *irrational*.

Proposition 2. *e is irrational.*

Proof. Suppose that e were rational. Then $e = a/b$ for some positive integers a and b . It follows that the number α defined by

$$\alpha = b! \left(e - 1 - \frac{1}{1!} - \frac{1}{2!} - \frac{1}{3!} - \cdots - \frac{1}{b!} \right)$$

is an integer (imagine multiplying the $b!$ through), and since e is defined by

$$e = \sum_{n=0}^{\infty} \frac{1}{n!},$$

α is positive. But the definition of e , along with the formula for the sum of a convergent geometric series, gives that

$$\begin{aligned} \alpha &= b! \left(\frac{1}{(b+1)!} + \frac{1}{(b+2)!} + \cdots \right) = \frac{1}{b+1} + \frac{1}{(b+1)(b+2)} + \cdots \\ &< \frac{1}{b+1} + \frac{1}{(b+1)^2} + \cdots = \frac{\frac{1}{b+1}}{1 - \frac{1}{b+1}} = \frac{1}{b} \leq 1. \end{aligned}$$

We find that $\alpha < 1$, a contradiction since α is a positive integer. \square

Note that e is an infinite sum of positive rational numbers; as such, it is the limit of an increasing sequence of rational numbers (namely, the sequence of partial sums), yet e itself is not rational. Indeed, \mathbb{R} has the property that every increasing sequence of rational numbers is either unbounded or converges to an element of \mathbb{R} . In fact, \mathbb{R} is the smallest such field, in the sense that any other field having this property contains \mathbb{R} .

We will see more irrational numbers later; indeed, it turns out that the irrationals are much more numerous than the rationals.

Another field that you may have worked with is the field of *complex numbers* \mathbb{C} :

$$\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R}\},$$

where $i = \sqrt{-1}$. Many of the commutative rings that we study in these notes are contained in \mathbb{C} .

DIVISIBILITY

Let us first focus on the integers \mathbb{Z} , and study their multiplicative structure. We begin by discussing how integers break down into simpler multiplicative parts.

Definition 3. If $a, b \in \mathbb{Z}$ and $b \neq 0$, we say that b divides a if there is an integer c such that $a = bc$.

We denote “ b divides a ” by $b \mid a$; synonyms for this that you may be familiar with are “ b is a divisor of a ”, “ b is a factor of a ”, “ a is a multiple of b ” and “ a is divisible by b ”. If b is not a divisor of a , we write $b \nmid a$.

Example 4. $3 \mid 12$, $7 \nmid 16$

Example 5. The positive divisors of 30 are 1, 2, 3, 5, 6, 10, 15 and 30.

Notice that any nonzero integer a is a divisor of 0 ($0 = a \cdot 0$) and is divisible by 1 ($a = 1 \cdot a$); a direct consequence of the former statement is the surprisingly useful

Corollary 6. *If a is an integer, and there is a positive integer b such that $b \nmid a$, then $a \neq 0$.*

We have only discussed divisibility in \mathbb{Z} ; notice that the notion of divisibility is trivial in \mathbb{Q} (i.e. if b is a nonzero rational number, then b divides every rational number) since we can invert any nonzero element of \mathbb{Q} . Indeed, divisibility is a trivial notion in every field. Therefore, when we discuss divisibility, we will mean it in the context of the integers, unless otherwise stated.

Proposition 7. *Let $a, b, c \in \mathbb{Z}$.*

- (1) *If $a \mid b$ and $b \mid c$, then $a \mid c$.*
- (2) *If $a \mid b$ and $a \mid c$, then for any integers x and y , $a \mid (bx + cy)$.*

Proof. (1) Since $a \mid b$ and $b \mid c$, there are integers m and n such that $b = ma$ and $c = nb$. Then $c = n(ma) = (nm)a$, so $a \mid c$.

(2) Since $a \mid b$ and $a \mid c$, there are integers m and n such that $b = ma$ and $c = na$. Then

$$bx + cy = (ma)x + (na)y = (mx + ny)a,$$

so $a \mid (bx + cy)$. □

THE PRIMES

Notice that every integer $a > 1$ has at least two positive divisors, namely 1 and a ; these are sometimes called the trivial divisors of a . If $d \mid a$ and $1 < d < a$, d is called a *proper* divisor of a .

Definition 8. An integer $p > 1$ is called *prime* if its only positive divisors are 1 and p (i.e., if it has no proper divisors). An integer $n > 1$ that is not prime is called *composite*.

Example 9. The first five prime numbers are 2, 3, 5, 7 and 11.

Prime numbers can be thought of, then, as multiplicatively the simplest positive integers. We now demonstrate their central place in multiplicative number theory.

Proposition 10. *If a positive integer n is composite, then the smallest proper divisor of n is prime.*

Proof. Let d be the smallest proper divisor of n . If d had a proper divisor m , then by Proposition 7 (1), m would be a proper divisor of n . But $m < d$, so we have a contradiction. Therefore d has no proper divisors, i.e., d is prime. \square

Theorem 11. *Every integer $n > 1$ is a product of prime numbers.*

Proof. By induction. Since 2 is prime, the theorem is certainly true for $n = 2$. Now suppose it is true for all the integers from 2 up to n . If $n + 1$ is prime, the theorem is true for $n + 1$. If $n + 1$ is composite, then by Proposition 10, it has a prime divisor p . Writing $n + 1 = mp$, we see that $2 \leq m \leq n$. By the induction hypothesis m is a product of primes, and therefore so is $n + 1 = mp$. \square

Example 12. $84 = 2 \cdot 42 = 2 \cdot 2 \cdot 21 = 2 \cdot 2 \cdot 3 \cdot 7$

We see that the primes are the multiplicative building blocks of \mathbb{Z} , and therefore it is natural to study them as a distinguished set. One obvious question to ask is “how many primes are there?”

Theorem 13. *(Euclid) There are infinitely many prime numbers.*

Proof. Let S be any finite set of prime numbers. Consider the integer

$$n = 1 + \prod_{p \in S} p.$$

If n is prime, then $n \notin S$, since n is larger than any element of S . If n is composite, then by Proposition 10, n has a prime factor q . Notice

that $q \notin S$, for if it were in S , then it would divide $n - 1 = \prod_{p \in S} p$, and then by Proposition 7 it would divide $n - (n - 1) = 1$, a contradiction. In any case, S does not contain all of the primes. \square

Now that we have shown that there are infinitely many primes, let us consider the question of identifying the primes among all the positive integers. Suppose we start from the very definition of a prime number: an integer $n > 1$ with no proper divisors. We can immediately see a foolproof way to determine whether a positive integer n is prime: search for proper divisors of n ; if we find one, n is not prime, and if we do not find any, n is prime. Since a proper divisor d of n satisfies $1 < d < n$, we know that this algorithm will involve no more than $n - 2$ steps.

Now that we have an algorithm, let us consider how to make it more efficient. First of all, let us note the obvious fact that the algorithm need not involve $n - 2$ steps for every n . Indeed, if we find a proper divisor of n , we may terminate our algorithm immediately at that point and conclude that n is not prime. In the extreme case that n is composite and the first integer we check happens to be a divisor of n , our algorithm consists of just one step. In general, then, the efficiency of our algorithm will depend partly upon our testing the integers that are most likely to be divisors of n first. Fortunately this is not difficult to do, because we know that one half of the positive integers are divisible by 2, one third of the positive integers are divisible by 3, and so on. Therefore we should test the integers for divisors of n in increasing order.

Notice that if n is not prime and we search in this way, we will automatically find the smallest proper divisor d of n first. Recall that, by Proposition 10, d is prime. We now prove another very important property of d .

Proposition 14. *If n is composite and d is the smallest proper divisor of n , then $d \leq \sqrt{n}$.*

Proof. Consider the alternative. \square

Combining Propositions 10 and 14, we obtain the following

Algorithm 15. (for determining whether n is prime) Search for divisors of n among the primes $\leq \sqrt{n}$, in increasing order beginning with 2. If a divisor is found, we conclude that n is composite and stop the search. If a divisor is not found, we conclude that n is prime.

Example 16. 113 is prime, because $\sqrt{113} = 10.63\dots$ and 113 is not divisible by 2, 3, 5 or 7.

Not a bad test; we don't need to check all the integers up to n for divisors, just those up to \sqrt{n} , and among these we only need to consider the primes.

There is something about our primality test that may bother you: to use it to test the primality of n , we need to have a list of the primes $\leq \sqrt{n}$. How do we find such a list? It turns out that we have an efficient way to find one, and for this we may thank the ancient Greek Eratosthenes. This process involves many of the ideas we used in deriving Algorithm 15.

Algorithm 17. (Sieve of Eratosthenes, for finding all the primes $\leq m$) List the integers from 2 to m , then apply the following iterative procedure to this list. The integers not eliminated in this process are the primes $\leq m$.

- (1) Determine the smallest integer p in the list.
- (2) If $p > \sqrt{m}$, stop. If $p \leq \sqrt{m}$, circle it and eliminate all multiples of p , except for p itself, from the list and go back to step (1).

Proof that Algorithm 17 works. First, it is clear that in the algorithm only composite numbers are eliminated. Let us now show that the circled numbers are all primes. We do this by induction. Clearly 2 is the smallest circled integer, and it is prime. Now suppose that the first k circled integers are primes. The sieve removed all of the multiples of these k primes (except themselves); hence the $(k + 1)$ st circled integer is not divisible by any of the primes that are smaller than it, and is therefore prime itself by Proposition 10. Finally, what about the remaining uncircled numbers? They remain because they are not multiples of any of the circled primes, which as we have seen are all the primes $\leq \sqrt{m}$; by Propositions 10 and 14, then, the remaining numbers are also prime.

COMMON DIVISORS

Consider the following problem: we wish to tile a rectangular floor that is 12 feet by 18 feet. For ease of cutting, we wish to use tiles that are square, of uniform size, and have integral side length s (in feet). Given these constraints, we would like to minimize the number of tiles we use (so as to minimize the amount of cutting). How shall we do this? Since we are using tiles of uniform size, it is clear that this is equivalent to finding the largest usable tile. If we use m rows and n columns of tiles, then, we have the following relationships:

$$ms = 18 \quad \text{and} \quad ns = 12.$$

These imply that s is a divisor of both 18 and 12, so the maximal s is the largest integer that is a divisor of both 18 and 12, namely 6.

Definition 18. If m and n are nonnegative integers, d is a *common divisor* of m and n if $d \mid m$ and $d \mid n$. If m and n are not both zero, the greatest common divisor (gcd) of m and n is denoted by (m, n) .

Example 19. The positive divisors of 20 are $\{1, 2, 4, 5, 10, 20\}$, and those of 35 are $\{1, 5, 7, 35\}$. The common divisors of 20 and 35 are $\{1, 5\}$, and so $(20, 35) = 5$.

Example 20. If m is a positive integer, then $(m, 0) = m$, since m is the largest divisor of m and any positive integer divides zero.

We see that computing the gcd of two positive integers can always be done in a straightforward way: list the divisors of each, and find the largest integer that appears in both lists. For large numbers, however, this procedure can become quite unwieldy. For example, 2310 has 32 divisors, and 1092 has 24 divisors; we might hope that there is a quicker way to find $(2310, 1092)$. Indeed, Euclid discovered an algorithm for doing this which has not been significantly improved to this date.

Algorithm 21. (Division Algorithm) Let m and n be positive integers. Then there exist unique integers q and r such that

$$n = mq + r$$

and $0 \leq r < m$.

Proof. Let $q = \lfloor n/m \rfloor$ (recall that if $r \in \mathbb{R}$, $\lfloor r \rfloor$ denotes the largest integer that is less than or equal to r). Since

$$(n/m) - 1 < \lfloor n/m \rfloor \leq n/m,$$

it follows that

$$0 = n - m(n/m) \leq n - mq < n - m((n/m) - 1) = m,$$

so let $r = n - mq$.

For uniqueness, suppose

$$n = mq_1 + r_1 = mq_2 + r_2$$

where $0 \leq r_1, r_2 < m$. Without loss we may assume that $r_1 \leq r_2$. Subtracting our two expressions for n , we find that

$$m(q_1 - q_2) = r_2 - r_1.$$

Hence $m \mid (r_2 - r_1)$. Since $0 \leq r_2 - r_1 < m$, it follows that $r_2 - r_1 = 0$, so $r_2 = r_1$. Then our last displayed equation gives $m(q_1 - q_2) = 0$, and since $m \neq 0$, it must be that $q_1 - q_2 = 0$, so $q_1 = q_2$. \square

Example 22. If we divide $m = 7$ into $n = 38$, we get a quotient of $q = 5$ and a remainder of $r = 3$.

Euclid's algorithm combines the division algorithm with the following

Proposition 23. *Let m, n and r be as in the division algorithm. Then $(m, n) = (m, r)$.*

Proof. By definition (m, n) divides m , and since $r = n - mq$, by Proposition 7 (2), (m, n) divides r . Since (m, n) is a common divisor of m and r , we have that $(m, n) \leq (m, r)$. On the other hand, (m, r) divides m , and since $n = mq + r$, Proposition 7 (2) tells us that (m, r) divides n . Hence (m, r) is a common divisor of m and n , and therefore $(m, r) \leq (m, n)$. We conclude that $(m, n) = (m, r)$. \square

Algorithm 24. (Euclidean Algorithm) Let m and n be positive integers with $m < n$. By the division algorithm, we have

$$n = mq_1 + r_1$$

with $0 \leq r_1 < m$. If $r_1 \neq 0$, we find by the division algorithm

$$m = r_1q_2 + r_2$$

with $0 \leq r_2 < r_1$. For $i \geq 2$, if $r_i \neq 0$, apply the division algorithm to obtain

$$r_{i-1} = r_iq_i + r_{i+1}$$

with $0 \leq r_{i+1} < r_i$. In this way we obtain a sequence

$$r_0 = m > r_1 > r_2 > r_3 > \cdots \geq 0.$$

Therefore $r_k = 0$ for some k . By Proposition 23, $(m, n) = (m, r_1) = (r_1, r_2) = \cdots = (r_{k-1}, r_k) = (r_{k-1}, 0) = r_{k-1}$.

Example 25. $(1092, 2310) = (1092, 126) = (126, 84) = (84, 42) = (42, 0) = 42$

The Euclidean Algorithm is extremely efficient; in fact, it allows us to find the gcd of two integers without factoring either one. As a byproduct of the algorithm, we can write (m, n) as a linear combination of m and n .

Proposition 26. *If m and n are positive integers, then there exist integers a and b such that*

$$(m, n) = am + bn.$$

Notice that, in the notation of Proposition 26, one of the integers a and b will be positive, and the other negative.

Example 27. Referring back to the previous example, the division that produced the gcd, 42, as remainder tells us that

$$42 = 126 - 84,$$

while the preceding divisions gave us

$$84 = 1092 - 8 \cdot 126 \quad \text{and} \quad 126 = 2310 - 2 \cdot 1092.$$

Then

$$\begin{aligned} 42 &= 126 - (1092 - 8 \cdot 126) = 9 \cdot 126 - 1092 = 9(2310 - 2 \cdot 1092) - 1092 \\ &= 9 \cdot 2310 - 19 \cdot 1092. \end{aligned}$$

Proposition 28. *If m and n are positive integers, then the set*

$$\{am + bn \mid a, b \in \mathbb{Z}\}$$

consists of all of the integer multiples of (m, n) .

Proof. By Proposition 7 (2), every element of the set is a multiple of (m, n) . And by Proposition 26, (m, n) is in the set. Therefore any multiple $d(m, n)$ of (m, n) is also in the set, for if we write $(m, n) = am + bn$, then

$$d(m, n) = d(am + bn) = (da)m + (db)n.$$

□

UNIQUE FACTORIZATION

We proved earlier (Theorem 11) that every positive integer is a product of primes. Our aim now is to show that the prime factorization of an integer is unique; that is, one always obtains the same prime factors no matter what path one takes to a prime factorization of a number:

$$90 = 3 \cdot 30 = 3 \cdot 3 \cdot 10 = 3 \cdot 3 \cdot 2 \cdot 5$$

$$90 = 5 \cdot 18 = 5 \cdot 2 \cdot 9 = 5 \cdot 2 \cdot 3 \cdot 3$$

We prove one preliminary result.

Proposition 29. *If p is prime and $p \mid mn$, then $p \mid m$ or $p \mid n$.*

Proof. Since $p \mid mn$, $mn = pq$. Suppose $p \nmid m$. Then $(m, p) = 1$, so by Proposition 26 we can find integers a and b such that $ma + pb = 1$. Then

$$n = n \cdot 1 = n(ma + pb) = (mn)a + npb = (pq)a + npb = (qa + nb)p,$$

and therefore $p \mid n$. \square

Theorem 30. *The prime factorization of an integer $n > 1$ is unique.*

Proof. Suppose n can be written as a product of primes as

$$n = p_1 \cdot p_2 \cdots p_k = q_1 \cdot q_2 \cdots q_\ell.$$

Then $p_1 \mid q_1(q_2 \cdots q_\ell)$, so by Proposition 29, either $p_1 \mid q_1$ (in which case $p_1 = q_1$), or $p_1 \mid q_2 \cdots q_\ell$. In the latter case, by the same argument either $p_1 = q_2$ or $p_1 \mid q_3 \cdots q_\ell$. By exhaustion we find that $p_1 = q_i$ for some $1 \leq i \leq \ell$. Canceling these, we find that

$$p_2 \cdots p_k = \prod_{j=1, j \neq i}^{\ell} q_j.$$

We may apply the same argument to show that $p_2 = q_j$ for some $j \neq i$; cancelling these and continuing in this fashion, we eventually eliminate all the p 's, so that the left hand product becomes 1. It follows that at this point, all of the q 's must have been canceled as well (else the product of the remaining q 's would be > 1). Thus the two factorizations are the same up to ordering. \square

Definition 31. If m is a positive integer and p is a prime, define $v_p(m)$ to be the highest power of p that divides m (this is a well-defined notion by Theorem 30).

Notice that

$$m = \prod_{p \text{ prime}} p^{v_p(m)},$$

and that $v_p(m) = 0$ for all but finitely many primes p .

We worked hard to prove that integers have unique factorizations into primes, a result that is probably quite familiar to you. It may be so familiar to you that the fact seems trivial. Is it? The answer to this question is no, in the following sense: there are number systems very similar to the integers in which factorization into primes is not unique.

Example 32. Consider the number system

$$\mathbb{Z}[\sqrt{-6}] = \{a + b\sqrt{-6} \mid a, b \in \mathbb{Z}\}$$

(in fact, this is a commutative ring; note that $\sqrt{-6} = \sqrt{6}i$ in our usual notation for \mathbb{C}). In this number system, 2 and 5 are “primes” in the sense that they have no nontrivial divisors (for example, the only divisors of 2 are ± 1 and ± 2); indeed, so are $2 + \sqrt{-6}$ and $2 - \sqrt{-6}$. The fact that

$$2 \cdot 5 = 10 = (2 + \sqrt{-6})(2 - \sqrt{-6})$$

shows that we do not have unique factorization in the number system $\mathbb{Z}[\sqrt{-6}]$, since 10 has two different factorizations.

Thinking back to the proof of Theorem 30, if we were to try the same technique to prove that $\mathbb{Z}[\sqrt{-6}]$ has unique factorization, the part that would fail would be the part involving Proposition 29, since the analog of this does not hold in $\mathbb{Z}[\sqrt{-6}]$. Our example shows this, for even though 2 has no nontrivial divisors in $\mathbb{Z}[\sqrt{-6}]$ and it is a divisor of the product $(2 - \sqrt{-6})(2 + \sqrt{-6})$, it is clearly not a divisor of either factor.

Unique factorization allows us to prove the irrationality of many real numbers.

Proposition 33. $\sqrt{7}$ is an irrational number.

Proof. Suppose $\sqrt{7}$ were rational. Then

$$\sqrt{7} = \frac{a}{b}$$

for some integers a and b . It follows that

$$a^2 = 7b^2.$$

Then $v_7(a^2) = v_7(7b^2)$, but $v_7(a^2) = 2v_7(a)$ is even and $v_7(7b^2) = 1 + 2v_7(b)$ is odd, a contradiction. \square

The proof of the preceding proposition generalizes easily to prove

Theorem 34. *If b and m are positive integers, and b not the m th power of another integer, then the m th root of b is irrational.*

Example 35. $4^{1/3}$, $\sqrt{27}$ are irrational.

CONGRUENCES

We develop here the language of congruences, which is extremely useful when discussing number theoretic questions.

Definition 36. Let a , b and m be integers, with $m > 0$. We say that a is *congruent to b modulo m* if $m \mid (a - b)$; in this case, we write $a \equiv b \pmod{m}$.

Example 37. $23 \equiv 8 \pmod{5}$, since $5 \mid (23 - 8) = 15$
 $53 \equiv -3 \pmod{8}$, since $8 \mid 53 - (-3) = 56$
 $28 \equiv 0 \pmod{7}$, since $7 \mid 28$

Example 38. The integers that are congruent to 0 modulo 3 are

$$\{\dots, -6, -3, 0, 3, 6, \dots\},$$

those congruent to 1 modulo 3 are

$$\{\dots, -5, -2, 1, 4, 7, \dots\},$$

and those congruent to 2 modulo 3 are

$$\{\dots, -4, -1, 2, 5, 8, \dots\}.$$

These sets are sometimes called the *congruence classes* (or *residue classes*) modulo 3. We see that every integer is congruent either to 0, 1 or 2 modulo 3; in general, modulo m , every integer is congruent to exactly one of $\{0, 1, 2, \dots, m - 1\}$. We often choose these m numbers as representatives of the congruence classes modulo m (of course, we could also choose other sets, such as $\{1, 2, \dots, m\}$). Note that the integers congruent to 0 modulo m are those that are multiples of m ; moreover, if a is a positive integer and we get a remainder of r upon dividing m into a , then $a \equiv r \pmod{m}$. When we work “modulo m ”, we consider integers a and b to be the same if $a \equiv b \pmod{m}$; hence the integers become a finite set, and are much easier to work with. We use this idea, for example, when specifying times. If it is 17 hours after midnight, we say that it is 5:00, not 17:00. This is because $17 \equiv 5 \pmod{12}$, i.e., we specify hours of the day modulo 12. Of course, with hours in military time we work modulo 24. In both cases, when dealing with minutes we work modulo 60. Notice that in standard time the representatives we choose for the hours (when we work modulo 12) are $1, 2, 3, \dots, 12$, while for the minutes (when we work modulo 60) we use the representatives $0, 1, 2, \dots, 59$.

We now show that we can do arithmetic “modulo m ”, and that this is consistent with the usual arithmetic in \mathbb{Z} .

Proposition 39. *Suppose a, b, c, d and m are integers, $m > 0$, $a \equiv c \pmod{m}$ and $b \equiv d \pmod{m}$. Then*

$$a + b \equiv c + d \pmod{m} \quad \text{and} \quad ab \equiv cd \pmod{m}.$$

Proof. Since $a \equiv c \pmod{m}$ and $b \equiv d \pmod{m}$, we know that $m \mid (a - c)$ and $m \mid (b - d)$. Then by Proposition 7 (2),

$$m \mid (a - c) + (b - d) = (a + b) - (c + d),$$

so $a + b \equiv c + d \pmod{m}$. Moreover, by Proposition 7 (1), $m \mid b(a - c) = ab - bc$ and $m \mid c(b - d) = bc - cd$. Hence Proposition 7 (2) gives that

$$m \mid (ab - bc) + (bc - cd) = ab - cd,$$

and therefore $ab \equiv cd \pmod{m}$. □

Notice that Proposition 39 implies that the congruence classes modulo m form a commutative ring under the addition and multiplication induced from \mathbb{Z} . This ring is denoted $\mathbb{Z}/m\mathbb{Z}$.

Example 40. Suppose we wish to know what

$$3294794857 \cdot 90983475983$$

is modulo 10. One way to do this is to multiply these numbers together and look at the result. Another way is to use Proposition 39; clearly

$$3294794857 \equiv 7 \pmod{10} \quad \text{and} \quad 90983475983 \equiv 3 \pmod{10},$$

and therefore

$$3294794857 \cdot 90983475983 \equiv 7 \cdot 3 \equiv 21 \equiv 1 \pmod{10}.$$

Example 41. Suppose we wish to calculate 4^{602} modulo 7. Rather than actually computing 4^{602} (which has hundreds of digits), we can simply note that

$$4^3 = 64 \equiv 1 \pmod{7},$$

and then by Proposition 39,

$$4^{602} = (4^3)^{200} \cdot 4^2 \equiv 1^{200} \cdot 4^2 \equiv 16 \equiv 2 \pmod{7}.$$

We see that “modular arithmetic” is often easier than the usual arithmetic in \mathbb{Z} , since there are only m congruence classes modulo m . Another way that modular arithmetic differs from that in \mathbb{Z} is that there are usually many units modulo m , whereas in \mathbb{Z} , the only units are -1 and 1 . For example, there is no integer a such that $5a = 1$, but modulo 7 we have that $5 \cdot 3 = 15 \equiv 1 \pmod{7}$.

Proposition 42. *Let a and m be integers, $m > 0$. There exists an integer b such that $ab \equiv 1 \pmod{m}$ if and only if $(a, m) = 1$.*

Proof. Suppose that $(a, m) = 1$. By Proposition 26, there exist integers b and n such that $ab + mn = 1$. Then $ab = 1 - mn \equiv 1 \pmod{m}$.

Now suppose that there exists an integer b such that $ab \equiv 1 \pmod{m}$. Then $ab - 1 = km$ for some integer k . By Proposition 7 (2), $(a, m) \mid ab - km = 1$, and therefore $(a, m) = 1$. \square

Corollary 43. *If $ax \equiv ay \pmod{m}$ and $(a, m) = 1$, then $x \equiv y \pmod{m}$.*

Proof. Since $(a, m) = 1$, we can find b such that $ab \equiv 1 \pmod{m}$. Then

$$x \equiv (ab)x \equiv b(ax) \equiv b(ay) \equiv (ab)y \equiv y \pmod{m}.$$

\square

Thus we may cancel a common factor a from a congruence modulo m if $(a, m) = 1$. Notice that we may NOT cancel in general if $(a, m) > 1$; for example,

$$6 \cdot 8 \equiv 6 \cdot 3 \pmod{15} \quad \text{even though} \quad 8 \not\equiv 3 \pmod{15}.$$

Definition 44. If a and m are integers, we say that a and m are *coprime*, or *relatively prime*, if $(a, m) = 1$.

We have seen that there are only m distinct congruence classes modulo m . Let us now study this ring further.

Definition 45. For $m \geq 1$, let

$$U(m) = \{1 \leq a \leq m \mid (a, m) = 1\}.$$

The *Euler phi function* ϕ is defined by $\phi(m) = \#U(m)$.

Notice that, by Proposition 42, the numbers in $U(m)$ are representatives for the units in the integers modulo m .

Example 46. $U(6) = \{1, 5\}$, $U(11) = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $U(12) = \{1, 5, 7, 11\}$, so $\phi(6) = 2$, $\phi(11) = 10$, $\phi(12) = 4$

Proposition 47. *If p is prime, then $\phi(p) = p - 1$.*

Proof. Since the only positive divisors of p are 1 and p ,

$$(a, p) = \begin{cases} 1 & \text{if } p \nmid a \\ p & \text{if } p \mid a \end{cases}$$

Therefore $U(p) = \{1, 2, \dots, p - 1\}$. □

The units in the integers modulo m have many interesting properties. Before we get to these, we prove two preliminary propositions.

Proposition 48. *If a, b and m are positive integers and $a \equiv b \pmod{m}$, then $(a, m) = (b, m)$.*

Proof. Since $a \equiv b \pmod{m}$, $a - b = km$ for some integer k . By definition, (b, m) divides m , and by Proposition 7 (2), (b, m) divides $a = b + km$. Since (b, m) is a common divisor of m and a , $(b, m) \leq (a, m)$. On the other hand, (a, m) is a divisor of m by definition, and Proposition 7 (2) tells us that (a, m) divides $b = a - km$. Since (a, m) is a common divisor of m and b , $(a, m) \leq (b, m)$. □

Proposition 49. *Let a, b and c be positive integers.*

- (1) *If $(a, b) = (a, c) = 1$, then $(a, bc) = 1$.*
- (2) *If $a \mid c$ and $b \mid c$ and $(a, b) = 1$, then $ab \mid c$.*
- (3) *If $a \mid bc$ and $(a, b) = 1$, then $a \mid c$.*

Proof. (1) By contradiction. Suppose $(a, bc) = d > 1$. Then d has a prime divisor p by Theorem 11. Since $p \mid d$ and $d \mid bc$, $p \mid bc$ by Proposition 7 (1). Then $p \mid b$ or $p \mid c$ by Proposition 29, so $p \mid (a, b)$ or $p \mid (a, c)$, a contradiction.

(2) Since $a \mid c$, $c = ka$ for some integer k . Moreover, by Proposition 26, we can find integers m and n such that $ma + nb = 1$. Then

$$k = k \cdot 1 = k(ma + nb) = cm + knb.$$

It follows by Proposition 7 (2) that $b \mid k$, so there is an integer ℓ such that $k = b\ell$. Hence $c = ka = \ell(ab)$, and $ab \mid c$.

(3) Since $a \mid bc$ and $(a, b) = 1$, there exist integers k, m and n such that $bc = ak$ and $ma + nb = 1$. Then

$$c = c(ma + nb) = cma + nbc = cma + nka = a(cm + nk),$$

and therefore $a \mid c$. □

Proposition 50. *If a and m are integers with $m \geq 1$ and $(a, m) = 1$, then there is an $r \leq \phi(m)$ such that $a^r \equiv 1 \pmod{m}$.*

Proof. Consider the integers

$$a, a^2, a^3, \dots, a^{\phi(m)+1}.$$

Since $(a, m) = 1$, $(a^n, m) = 1$ for any $n \geq 1$ by Proposition 49 (1). Hence the list above has $\phi(m) + 1$ elements, each coprime to m . By Proposition 48, the least residues of these are in the set $U(m)$. Since $U(m)$ has $\phi(m)$ elements, it follows that these powers of a cannot all be distinct modulo m . So $a^i \equiv a^j \pmod{m}$ for some $i < j$. By Corollary 43, then, $a^{j-i} \equiv 1 \pmod{m}$. □

Definition 51. If a and m are coprime integers with $m \geq 1$, we call the smallest positive integer r such that $a^r \equiv 1 \pmod{m}$ the *order* of a modulo m .

Example 52. The order of 3 modulo 11 is 5, since

$$3^1 \equiv 3 \pmod{11}, \quad 3^2 \equiv 9 \pmod{11}, \quad 3^3 \equiv 5 \pmod{11}$$

$$3^4 \equiv 4 \pmod{11}, \quad 3^5 \equiv 1 \pmod{11}.$$

Proposition 53. *Suppose a and m are integers with $m > 0$ and $(a, m) = 1$. If s is a positive integer such that $a^s \equiv 1 \pmod{m}$, then the order of a modulo m divides s .*

Proof. Denote by t the order of a modulo m . By the division algorithm we obtain

$$s = qt + r \quad \text{with} \quad 0 \leq r < t.$$

Then

$$1 \equiv a^s \equiv a^{qt+r} \equiv (a^t)^q \cdot a^r \equiv a^r \pmod{m}.$$

Since $r < t$, $r = 0$ by the definition of order, and therefore $t \mid s$. □

Theorem 54. (*Fermat*) *If $m \geq 1$ and $(a, m) = 1$, then $a^{\phi(m)} \equiv 1 \pmod{m}$.*

Proof. Consider the set of $\phi(m)$ congruence classes modulo m represented by the set of integers

$$U(m) = \{n \mid 1 \leq n \leq m \text{ and } (n, m) = 1\}.$$

By Corollary 43, the $\phi(m)$ integers

$$\{an \mid n \in U(m)\},$$

also represent $\phi(m)$ distinct congruence classes modulo m . Indeed, since $(an, m) = 1$ by Proposition 49 (1), it follows that the congruence classes represented by these two sets are the same. Therefore

$$\prod_{n \in U(m)} n \equiv \prod_{n \in U(m)} an \equiv a^{\phi(m)} \cdot \prod_{n \in S} n \pmod{m}.$$

Upon canceling, we find that $a^{\phi(m)} \equiv 1 \pmod{m}$. □

Corollary 55. *If $m \geq 1$ and $(a, m) = 1$, then the order of a modulo m divides $\phi(m)$.*

Example 56. Suppose we want to know the order of 5 modulo 257. Since 257 is prime and $5 \nmid 257$, we know that this order is a divisor of $\phi(257) = 256 = 2^8$. Since the divisors of 2^8 are $1, 2, 2^2, \dots, 2^8$, to find the order of 5 we need only compute

$$5^2 \equiv 25 \pmod{257}, \quad 5^4 = (5^2)^2 = 25^2 \equiv 111 \pmod{257}$$

$$5^8 \equiv 111^2 \equiv 242 \pmod{257}, \quad 5^{16} \equiv 242^2 \equiv 225 \pmod{257}$$

$$5^{32} \equiv 225^2 \equiv 253 \pmod{257}, \quad 5^{64} \equiv 253^2 \equiv 16 \pmod{257}$$

$$5^{128} \equiv 16^2 \equiv 256 \pmod{257},$$

and therefore 5 has order 256 modulo 257.

Corollary 57. *If p is prime and $p \nmid a$, then $a^{p-1} \equiv 1 \pmod{p}$.*

Corollary 58. *If p is prime and a is any integer, then $a^p \equiv a \pmod{p}$.*

We end this section by considering the question of simultaneous congruences.

Theorem 59. *(Chinese Remainder Theorem) Suppose m_1, m_2, \dots, m_r are positive integers such that $(m_i, m_j) = 1$ if $i \neq j$. Let a_1, a_2, \dots, a_r be any integers, and write $m = m_1 m_2 \cdots m_r$. Then there exists an integer x such that*

$$x \equiv a_i \pmod{m_i}$$

for every $1 \leq i \leq r$. Moreover, for any integer y satisfying these congruences, $y \equiv x \pmod{m}$.

Proof. For each $1 \leq j \leq r$, m/m_j is an integer, and by Proposition 49 (1), $(m/m_j, m_j) = 1$. Then by Proposition 42, there is an integer b_j such that $(m/m_j)b_j \equiv 1 \pmod{m_j}$. Let

$$x = \sum_{j=1}^r (m/m_j)b_j a_j.$$

Since $(m/m_j)b_j \equiv 0 \pmod{m_i}$ if $i \neq j$, it follows that for every $1 \leq i \leq r$,

$$x \equiv (m/m_i)b_i a_i \equiv a_i \pmod{m_i}.$$

Now, if y is any solution of the stated congruences, then $m_i \mid (x - y)$ for all i . By Proposition 49 (2), it follows that $m \mid (x - y)$, and thus $x \equiv y \pmod{m}$. □

POLYNOMIALS

Some of the most famous problems in number theory are stated in terms of polynomials. Perhaps the most renowned of all, and which remained unsolved for 350 years, is known as Fermat's Last Theorem.

Theorem 60. (Wiles) *If x, y and z are positive integers and $n \geq 3$, then $x^n + y^n \neq z^n$.*

There are many techniques for studying integer solutions of polynomial equations. We illustrate some here.

Example 61. The equation $x^2 + 5y^2 = -2$ has no integer solutions, because if $x, y \in \mathbb{Z}$, the *LHS* (left hand side) is nonnegative. Notice that this equation has no *real* solutions either, by the same argument.

Example 62. The equation $x^2 + 5y^2 = 2$ has no integer solutions. To see this, notice first that if $x, y \in \mathbb{Z}$, then $LHS \geq 5y^2 \geq 5$ whenever $y \neq 0$. So any solution would have $y = 0$, but this leaves $x^2 = 2$, which has no integer solution. Notice that this equation has infinitely many *real* solutions; indeed, its graph in the real plane is an ellipse.

Example 63. The equation $x^2 - 5y^2 = 2$ has no integer solutions. To see this, notice that if $x, y \in \mathbb{Z}$, then $LHS \equiv x^2 \pmod{5}$. Now,

$$\begin{aligned} 0^2 &\equiv 0 \pmod{5}, & 1^2 &\equiv 1 \pmod{5}, & 2^2 &\equiv 4 \pmod{5}, \\ 3^2 &\equiv 4 \pmod{5}, & 4^2 &\equiv 1 \pmod{5}, \end{aligned}$$

and since every integer is congruent to one of $\{0, 1, 2, 3, 4\}$ modulo 5, it follows that x^2 cannot be congruent to 2 modulo 5. Notice that this equation has infinitely many real solutions; indeed, its graph is a hyperbola.

Example 64. The equation $x^2 + 5y^2 = 1$ has exactly two integer solutions, $(x, y) = (\pm 1, 0)$. To see this, note that $LHS \geq 5y^2 \geq 5$ if $y \neq 0$, so any integer solution would have $y = 0$, leaving $x^2 = 1$.

Example 65. The equation $x^2 - 5y^2 = 1$ has infinitely many integer solutions. To see this, note first that $(x, y) = (9, 4)$ is an integer solution. Now, we may factor the *LHS*, giving us

$$(x - y\sqrt{5})(x + y\sqrt{5}) = 1,$$

and so $(9 - 4\sqrt{5})(9 + 4\sqrt{5}) = 1$. For each $n \geq 1$, define $a_n, b_n \in \mathbb{Z}$ by

$$(9 + 4\sqrt{5})^n = a_n + b_n\sqrt{5}.$$

Then $(9 - 4\sqrt{5})^n = a_n - b_n\sqrt{5}$, and since

$$1 = 1^n = [(9 + 4\sqrt{5})(9 - \sqrt{5})]^n = (a_n + b_n\sqrt{5})(a_n - b_n\sqrt{5}) = a_n^2 - 5b_n^2,$$

we find that $(x, y) = (a_n, b_n)$ is an integer solution of our equation for every $n \geq 1$. And since $a_{n+1} > a_n$ for every $n \geq 1$, it follows that these constitute infinitely many distinct integer solutions. For example,

$$(9 + 4\sqrt{5})^2 = 161 + 72\sqrt{5},$$

and so $(x, y) = (161, 72)$ is a solution.

Perhaps the most widely known polynomial equation is that arising in the Pythagorean Theorem:

$$x^2 + y^2 = z^2.$$

Let us now study this equation. Notice that if (x, y, z) is an integer solution of this equation (called a *Pythagorean triple*), then $(x/z, y/z)$ is a point with rational coordinates (a *rational point*) on the unit circle. We now use a geometric argument to show that we can parametrize the rational points on the unit circle (and therefore the Pythagorean triples).

Consider a line through the point $(-1, 0)$ with rational slope m . Its equation is $y = m(x + 1)$, and such a line intersects the unit circle in a second point. To see what this other point is, we substitute and factor: $1 = x^2 + [m(x + 1)]^2$, and therefore

$$0 = (m^2 + 1)x^2 + 2m^2x + (m^2 - 1) = (x + 1)[(m^2 + 1)x + (m^2 - 1)].$$

Therefore the second point of intersection is

$$(x, y) = \left(\frac{1 - m^2}{1 + m^2}, \frac{2m}{1 + m^2} \right).$$

Hence every $m \in \mathbb{Q}$ gives us a rational point on the unit circle. Moreover, if $(a, b) \neq (-1, 0)$ is a rational point on the unit circle, then the line through it and $(-1, 0)$ has slope $b/(a + 1)$, a rational number. It follows that using our lines of rational slope through $(-1, 0)$, we obtain all of the rational points on the unit circle.

Letting $m = a/b$ in our preceding discussion, we find:

Theorem 66. *Every Pythagorean triple (x, y, z) with y even has the form*

$$x = b^2 - a^2, \quad y = 2ab, \quad z = a^2 + b^2,$$

with $a, b \in \mathbb{Z}$.

Let us now focus on polynomials of one variable. Let $\mathbb{Z}[x]$ denote the set of polynomials in one variable x with coefficients in \mathbb{Z} . We consider quadratic polynomials first.

Proposition 67. *Let p be prime, $x \in \mathbb{Z}$. Then $x^2 \equiv 1 \pmod{p}$ if and only if $x \equiv \pm 1 \pmod{p}$.*

Proof. Using Proposition 29,

$$x^2 \equiv 1 \pmod{p} \iff p \mid (x^2 - 1) = (x - 1)(x + 1)$$

$$\iff p \mid (x - 1) \quad \text{or} \quad p \mid (x + 1)$$

$$\iff x \equiv 1 \pmod{p} \quad \text{or} \quad x \equiv -1 \pmod{p}$$

□

Theorem 68. (*Wilson's Theorem*) *If p is prime, then $(p - 1)! \equiv -1 \pmod{p}$.*

Proof. One can easily check that the theorem is true if $p = 2$ or $p = 3$, so let us assume that $p \geq 5$. For each $1 \leq a \leq p - 1$, $(a, p) = 1$, so by Proposition 42 there is a $b \in \mathbb{Z}$ such that $ab \equiv 1 \pmod{p}$. By Proposition 67, $b \equiv a \pmod{p}$ only if $a = 1$ or $a = p - 1$. Therefore, in the product $(p - 1)!$, we may pair each $2 \leq a \leq p - 2$ with its multiplicative inverse and multiply them (giving 1 modulo p). It follows that

$$(p - 1)! = 1 \cdot \prod_{a=2}^{p-2} a \cdot (p - 1) \equiv 1 \cdot 1 \cdot (p - 1) \equiv -1 \pmod{p}.$$

□

Example 69. If $p = 7$,

$$6! = 1 \cdot (2 \cdot 4) \cdot (3 \cdot 5) \cdot 6 \equiv 1 \cdot 1 \cdot 1 \cdot 6 \equiv -1 \pmod{7}.$$

Some very interesting phenomena arise in the study of polynomial equations modulo p .

Proposition 70. *Let p be a prime. Then $x^2 \equiv -1 \pmod{p}$ has an integer solution x if and only if $p = 2$ or $p \equiv 1 \pmod{4}$.*

Proof. If $p = 2$, take $x = 1$. Suppose p is odd from now on. If $p \equiv 1 \pmod{4}$, we now show that we can build a solution. Using Theorem 68 and the fact that $(p - 1)/2$ is even,

$$\begin{aligned} -1 &\equiv (p - 1)! = \left[1 \cdot 2 \cdots \frac{p - 1}{2} \right] \left[\frac{p + 1}{2} \cdots (p - 2) \cdot (p - 1) \right] \\ &= \prod_{j=1}^{\frac{p-1}{2}} j \cdot \prod_{j=1}^{\frac{p-1}{2}} (p - j) \equiv \prod_{j=1}^{\frac{p-1}{2}} j(p - j) \equiv \prod_{j=1}^{\frac{p-1}{2}} (-j^2) \\ &\equiv (-1)^{\frac{p-1}{2}} \prod_{j=1}^{\frac{p-1}{2}} j^2 \equiv \left(\prod_{j=1}^{\frac{p-1}{2}} j \right)^2 \pmod{p}, \end{aligned}$$

so indeed we have a solution if $p \equiv 1 \pmod{4}$. Now suppose that we have a solution x . Clearly $p \nmid x$, and therefore

$$(-1)^{\frac{p-1}{2}} \equiv (x^2)^{\frac{p-1}{2}} \equiv x^{p-1} \equiv 1 \pmod{p}$$

by Corollary 57. Since -1 is not congruent to 1 modulo p , it follows that $(p - 1)/2$ is even, and therefore $p \equiv 1 \pmod{4}$. □

We have just seen that the question of whether $x^2 \equiv -1 \pmod{p}$ has a solution naturally divides the odd primes up into congruence classes (1 and 3) modulo 4. Notice that Proposition 70 implies that the polynomial $x^2 + 1$ is *never* equal to a prime that is congruent to 3 modulo 4 when x is an integer. An interesting open question in number theory is: Is $x^2 + 1$ prime for infinitely many integer values of x ? Indeed, there is no polynomial in $\mathbb{Z}[x]$ of degree at least two which has been proven to be prime for infinitely many integer values of x .

We now expand our view to polynomials of higher degree.

Algorithm 71. (Division Algorithm for polynomials) Given $f(x), g(x) \in \mathbb{Z}[x]$ with $f(x)$ monic and $\deg(f(x)) \geq 1$, there exist unique $q(x), r(x) \in \mathbb{Z}[x]$ such that

$$g(x) = q(x)f(x) + r(x)$$

with $1 \leq \deg(r(x)) < \deg(f(x))$.

Proof. For the existence of $q(x)$ and $r(x)$, consider the set

$$S = \{g(x) - h(x)f(x) \mid h(x) \in \mathbb{Z}[x]\}.$$

Since f has positive degree, the set S clearly has elements of positive degree. Let $r(x)$ be an element of S of minimal degree. Then $r(x) = g(x) - q(x)f(x)$ for some $q(x) \in \mathbb{Z}[x]$. We claim that $0 \leq \deg(r(x)) < \deg(f(x))$, because if $\deg(r(x)) \geq \deg(f(x))$, then if the leading coefficient of $r(x)$ is c , we have that

$$g(x) - [q(x) + c \cdot x^{\deg(r(x)) - \deg(f(x))}] f(x)$$

is an element of S with degree less than $\deg(r(x))$, a contradiction.

For uniqueness, suppose

$$g(x) = q_1(x)f(x) + r_1(x) = q_2(x)f(x) + r_2(x)$$

with $0 \leq \deg(r_1(x)) \leq \deg(r_2(x)) < \deg(f(x))$. Then $r_1(x) - r_2(x) = (q_2(x) - q_1(x))f(x)$. Since $\deg(r_1(x) - r_2(x)) < \deg(f(x))$, it must be that $q_2(x) - q_1(x) = 0$. Thus $q_1(x) = q_2(x)$; this along with the last displayed equation implies that $r_1(x) = r_2(x)$. □

Theorem 72. *If $f \in \mathbb{Z}[x]$ is a monic polynomial and $1 \leq \deg(f(x)) < p$, then f has no more than $\deg(f(x))$ distinct roots modulo p .*

Proof. By induction on the degree of f . If $\deg(f(x)) = 1$, then $f(x) = x - a$ for some $a \in \mathbb{Z}$. Clearly this has exactly one root: $x \equiv a \pmod{p}$. Now suppose that the theorem is true for all polynomials of the given form of degree $\leq n$. If $n = p - 1$ we're done. If not, let f be a monic polynomial of degree $n + 1$. If f has no roots modulo p , the theorem is certainly true for f . So let us assume that there is an integer a such that $f(a) \equiv 0 \pmod{p}$. By the Division Algorithm,

$$f(x) = (x - a)q(x) + r(x)$$

with $0 \leq \deg(r(x)) < 1$. Hence $r(x)$ has degree zero, i.e. $r(x) = r \in \mathbb{Z}$. Since $0 \equiv f(a) \equiv r \pmod{p}$, we have that $p \mid r$. Moreover, since $\deg(q(x)) < \deg(f(x))$, we have by the induction hypothesis that $q(x)$ has $< \deg(f(x))$ roots modulo p . Now, for any $b \in \mathbb{Z}$,

$$f(b) = (b - a)q(b) + r \equiv (b - a)q(b) \pmod{p},$$

so by Proposition 29,

$$f(b) \equiv 0 \pmod{p} \iff b \equiv a \pmod{p} \quad \text{or} \quad q(b) \equiv 0 \pmod{p}.$$

Thus f has $\leq \deg(f(x))$ solutions modulo p . □

Notice that by Corollary 58, when we are working modulo p , there is no need to consider polynomials of degree $\geq p$.

Example 73. The assumption that p is prime is crucial in Theorem 72. For example, $x^2 \equiv 1 \pmod{8}$ has four solutions modulo 8: 1, 3, 5 and 7. Moreover, $x^2 \equiv 1 \pmod{24}$ has 8 solutions modulo 24: 1, 5, 7, 11, 13, 17, 19 and 23.

Proposition 74. *Suppose p is an odd prime, $a \in \mathbb{Z}$ and $p \nmid a$. Then $x^2 \equiv a \pmod{p}$ has zero or two solutions modulo p .*

Proof. If there are no solutions, we are done. On the other hand, if there exists an $x \in \mathbb{Z}$ such that $x^2 \equiv a \pmod{p}$, then $-x$ is also a solution, and since p is odd, x and $-x$ are not congruent modulo p . Thus we have at least two solutions modulo p in the latter case, and by Theorem 72, we have exactly two solutions modulo p . \square

We now endeavor to use our results on polynomials to describe the structure of the set of units of $\mathbb{Z}/p\mathbb{Z}$ when p is a prime. Recall that, by Proposition 42, the congruence classes represented by $\{1, 2, \dots, p-1\}$ are the units of $\mathbb{Z}/p\mathbb{Z}$.

Proposition 75. *Suppose a, b and m are integers with $m > 0$ and $(a, m) = (b, m) = 1$. If a has order k modulo m , b has order ℓ modulo m and $(k, \ell) = 1$, then ab has order $k\ell$ modulo m .*

Proof. By Proposition 49 (1), $(ab, m) = 1$, so ab has an order modulo m ; denote this order by r . Now,

$$(ab)^{k\ell} = (a^k)^\ell (b^\ell)^k \equiv 1^\ell \cdot 1^k \equiv 1 \pmod{m},$$

so by Proposition 53, $r \mid k\ell$. Furthermore,

$$b^{kr} = 1 \cdot b^{kr} = (a^k)^r \cdot (b^{kr}) = (ab)^{kr} = ((ab)^r)^k \equiv 1^k \equiv 1 \pmod{m},$$

so by Proposition 53, $\ell \mid kr$. Since $(k, \ell) = 1$, it follows by Proposition 49 (3) that $\ell \mid r$. By considering $a^{\ell r}$, an exactly analogous argument shows that $k \mid r$. Since $(k, \ell) = 1$, Proposition 49 (2) tells us that $k\ell \mid r$. As we showed earlier that $r \mid k\ell$, it must be that $r = k\ell$. \square

Definition 76. If a and m are integers with $m > 0$ and $(a, m) = 1$, we call a a *primitive root modulo m* if a has order $\phi(m)$ modulo m .

Notice that by Theorem 54, $\phi(m)$ is the maximal order that an element can have modulo m .

Example 77. 3 is a primitive root modulo 7, since it has order 6:

$$3^1 \equiv 3 \pmod{7}, \quad 3^2 \equiv 2 \pmod{7}, \quad 3^3 \equiv 6 \pmod{7},$$

$$3^4 \equiv 4 \pmod{7}, \quad 3^5 \equiv 5 \pmod{7}, \quad 3^6 \equiv 1 \pmod{7}.$$

In general, if a is a primitive root modulo a prime p , all of the congruence classes $\{1, 2, \dots, p-1\}$ are powers of a , i.e., the positive powers of a run through all of the nonzero congruence classes modulo p .

Theorem 78. *If p is prime, then there exists a primitive root modulo p .*

Proof. If $p = 2$, then $x = 1$ is a primitive root modulo p . Assume from now on that p is odd. We claim first that if $d \mid (p-1)$ and $d > 0$, then $x^d \equiv 1 \pmod{p}$ has exactly d solutions modulo p . To see this, write $p-1 = de$. We know that

$$y^e - 1 = (y-1)(y^{e-1} + y^{e-2} + \dots + y + 1),$$

Letting $y = x^d$, it follows that

$$x^{p-1} - 1 = x^{de} - 1 = (x^d - 1)(x^{d(e-1)} + x^{d(e-2)} + \dots + x^d + 1).$$

By Corollary 57, $x^{p-1} - 1$ has $p-1$ roots modulo p . And by Theorem 72, $x^{d(e-1)} + \dots + 1$ has $\leq d(e-1) = (p-1) - d$ roots. By Proposition 29, it follows that $x^d - 1$ has at least d roots modulo p , and by Theorem 72 it therefore has exactly d roots modulo p .

For each prime divisor q of $p-1$, define $\alpha(q) := q^{v_q(p-1)}$. By what we have just shown, $x^{\alpha(q)} - 1$ has $\alpha(q)$ roots modulo p . Moreover, $x^{\alpha(q)/q} - 1$ has $\alpha(q)/q$ roots modulo p , each of which, of course, is a root of $x^{\alpha(q)} - 1$. Thus $x^{\alpha(q)} - 1$ has $\alpha(q) - (\alpha(q)/q)$ roots that are not roots of $x^{\alpha(q)/q} - 1$. By Proposition 53, the roots of $x^d - 1 \equiv 0 \pmod{p}$ are those elements whose order modulo p is a divisor of d . Thus we have shown the existence of $\alpha(q) - (\alpha(q)/q)$ elements of order $\alpha(q)$.

To finish the proof, for each prime divisor q of $p-1$, let β_q be an integer whose order modulo p is $\alpha(q)$. Then by Proposition 75, $\prod_{q \mid (p-1)} \beta_q$ has order $\prod_{q \mid (p-1)} \alpha(q) = p-1$. \square

PSEUDOPRIMES

Recall that we have an algorithm for testing whether a positive integer is prime (Algorithm 15). Since we have a primality test, our goal should now be to find the most efficient primality test possible. Notice that Corollary 57 gives a necessary condition for a number n to be prime: if n is prime, it must be the case that $a^{n-1} \equiv 1 \pmod{n}$ for any integer a such that $n \nmid a$. Therefore we can use Corollary 57 to prove that numbers are composite. For example, suppose we wanted to prove that 91 is composite. One way would be to exhibit a proper divisor of 91, for example 7. Another way to prove this would be to note that $2 \nmid 91$, and that $2^{90} \equiv 64 \pmod{91}$; it follows from Corollary 57 that 91 is not prime.

This second method of proof that 91 is composite might seem a bit strange and indirect. However, as we will see, its indirectness is its strength; indeed, notice that it allowed us to prove that 91 is composite WITHOUT FINDING A PROPER DIVISOR OF 91.

Of course, we could take this same approach for any odd integer n :

$$2^8 \equiv 4 \pmod{9}, \quad 2^{14} \equiv 4 \pmod{15}, \quad 2^{20} \equiv 4 \pmod{21},$$

$$2^{24} \equiv 16 \pmod{25}, \quad 2^{26} \equiv 13 \pmod{27}, \quad 2^{32} \equiv 4 \pmod{33}.$$

Notice that our “2-test” detects every odd composite ≤ 33 . This is very intriguing because, as we will now see, applying the “2-test” takes much less work than using Algorithm 15.

Definition 79. For any real number x , we denote by $\pi(x)$ the number of primes that are less than or equal to x .

Example 80. The first 8 primes are 2, 3, 5, 7, 11, 13, 17 and 19. Therefore $\pi(10) = 4$, $\pi(13) = 6$, $\pi(14) = 6$ and $\pi(19) = 8$.

In the “worst case”, which is when n is prime, Algorithm 15 involves testing all of the primes up to \sqrt{n} for divisors; that is, it will take $\pi(\sqrt{n})$ steps. How about our “2-test”? We may implement this test in an efficient way as follows. If $n - 1 = \sum_{j=1}^{\lfloor \log_2(n-1) \rfloor} a_j 2^j$ ($a_j = 0$ or 1) is the binary expansion of $n - 1$, we may by successive squaring compute

$$2^2, \quad 2^4 = (2^2)^2, \quad 2^8 = (2^4)^2, \dots, 2^{\lfloor \log_2(n-1) \rfloor} \pmod{n}.$$

Then

$$2^{n-1} \equiv \prod_{j \text{ with } a_j=1} 2^{2^j} \pmod{n}.$$

This entire process involves no more than $2\lfloor \log_2(n) \rfloor$ multiplications. If we compare $\pi(\sqrt{n})$ to $2\lfloor \log_2(n) \rfloor$ for increasingly large values of n , we see that the “2-test” becomes much more efficient than Algorithm 15 as n grows:

$$\pi(\sqrt{10000}) = 25, \quad 2\lfloor \log_2(10000) \rfloor = 26.57\dots$$

$$\pi(\sqrt{1000000}) = 168, \quad 2\lfloor \log_2(1000000) \rfloor = 39.86\dots$$

$$\pi(\sqrt{100000000}) = 1229, \quad 2\lfloor \log_2(100000000) \rfloor = 53.15\dots$$

The question, then, is “Is the 2-test a primality test?” Unfortunately the answer is “no”; indeed, the smallest composite number that “passes” the 2-test is $341 = 11 \cdot 31$. Thus the 2-test is not a primality test.

You might be thinking “what is so special about 2?” The answer is “nothing”. Indeed, for any n such that $3 \nmid n$ we may subject n to a “3-test” as well to try to determine whether n is composite. Indeed, if we apply it to 341 we find that

$$3^{340} \equiv 56 \pmod{341},$$

and we may conclude by Theorem 54 that 341 is composite.

Like the 2-test, the 3-test is not a primality test either; the smallest composite that passes the 3-test is 91. However, we saw that the composite 341, which passes the 2-test, does not pass the 3-test; also, the composite 91 passes the 3-test, but does not pass the 2-test. Perhaps the 2-test and 3-test together constitute a primality test?

Unfortunately this is not the case either. The smallest composite that passes both the 2-test and the 3-test is 1105. So we have not found another primality test yet, but perhaps we are simply not working hard enough yet. Indeed, recall that for a positive integer n , Theorem 54 offers us an “ a -test” for any integer a such that $(a, n) = 1$ (of course, since these tests involve computations modulo n , we may as well assume that $1 \leq a \leq n$). And since each a -test is much more efficient (for large n) than Algorithm 15, we could apply multiple a -tests and still do less work than if we were to apply Algorithm 15.

Definition 81. If n is composite and $(a, n) = 1$, we call n a *pseudoprime to the base a* if $a^{n-1} \equiv 1 \pmod{n}$.

Example 82. As we discussed above, 341 is a pseudoprime to the base 2, and 91 is a pseudoprime to the base 3.

Example 83. {341, 561, 645, 1105, 1387, 1729, 1905} is the set of all pseudoprimes to the base 2 that are ≤ 2000 , and

$$\{91, 121, 286, 671, 703, 949, 1105, 1541, 1729, 1891\}$$

is the set of all pseudoprimes to the base 3 that are ≤ 2000 , and

$$\{4, 124, 217, 561, 781, 1541, 1729, 1891\}$$

is the set of all pseudoprimes to the base 5 that are ≤ 2000 .

We see from the last example that the only composite less than 2000 that passes 2-test, 3-test and 5-test is 1729. We might hope that finding a primality test simply comes down to applying enough a -tests.

Definition 84. A composite integer n is called a *Carmichael number* if it is a pseudoprime to the base a for every integer a with $(a, n) = 1$.

The question of whether a battery of a -tests will constitute a primality test comes down to the question “are there any Carmichael numbers, and if so, how many?”

For our goal of using Theorem 54 to create a new primality test, this question has the worst possible answer; there DO exist Carmichael numbers (561 is the smallest), and there are infinitely many of them.

QUADRATIC RESIDUES

We have seen some of the interesting aspects of quadratic polynomials. We now develop language that will make discussion of this topic easier.

Definition 85. Let m be a positive integer. If a is an integer that is coprime to m , we say that a is a *quadratic residue modulo m* if there exists an integer x such that $x^2 \equiv a \pmod{m}$; if no such x exists, we call a a *quadratic nonresidue modulo m* .

Notice that by the very definition of quadratic residue, if a and b are integers such that $(a, m) = 1 = (b, m)$ and $a \equiv b \pmod{m}$, then a is a quadratic residue modulo m if and only if b is a quadratic residue modulo m .

Example 86. Since $1^2 \equiv 1 \pmod{5}$, $2^2 \equiv 4 \pmod{5}$, $3^2 \equiv 4 \pmod{5}$ and $4^2 \equiv 1 \pmod{5}$, 1 and 4 are quadratic residues modulo 5, and 2 and 3 are quadratic nonresidues modulo 5.

Example 87. If p is prime, by Proposition 70, -1 is a quadratic residue modulo p if and only if $p = 2$ or $p \equiv 1 \pmod{4}$.

Proposition 88. *Let p be an odd prime. Then exactly $(p - 1)/2$ of the integers $\{1, 2, \dots, p - 1\}$ (i.e., exactly half of them) are quadratic residues.*

Proof. Follows directly from Proposition 74. □

Definition 89. If p is an odd prime and a is an integer, we define the *Legendre symbol* $\left(\frac{a}{p}\right)$ by

$$\left(\frac{a}{p}\right) = \begin{cases} 1 & \text{if } a \text{ is a quadratic residue modulo } p, \\ -1 & \text{if } a \text{ is a quadratic nonresidue modulo } p, \\ 0 & \text{if } p \mid a. \end{cases}$$

Notice that, by the definition of quadratic residue, if $a \equiv b \pmod{p}$, then $\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$.

Proposition 90. *Let p be an odd prime, a an integer such that $p \nmid a$, and suppose g is a primitive root modulo p . Then $\left(\frac{a}{p}\right) = 1$ if and only if $g^{2n} \equiv a \pmod{p}$ for some positive integer n .*

Proof. First suppose that $\left(\frac{a}{p}\right) = 1$. Then there exists an integer x such that $x^2 \equiv a \pmod{p}$. Notice that since $p \nmid a$, $p \nmid x$. Hence $x \equiv g^n \pmod{p}$ for some positive integer n , and therefore

$$a \equiv x^2 \equiv (g^n)^2 \equiv g^{2n} \pmod{p}.$$

Now suppose that $a \equiv g^{2n}$ for some integer n . Then $a \equiv (g^n)^2 \pmod{p}$, and therefore $\left(\frac{a}{p}\right) = 1$. \square

Recall that Proposition 70 tells us that if p is an odd prime, then $\left(\frac{-1}{p}\right) = 1$ if and only if $p \equiv 1 \pmod{4}$. Since $(p-1)/2$ is even if and only if $p \equiv 1 \pmod{4}$, we have that

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}}.$$

Proposition 91. *If p is an odd prime, a is an integer and $p \nmid a$, then*

$$\left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \pmod{p}.$$

Proof. By Corollary 57, $(a^{(p-1)/2})^2 \equiv 1 \pmod{p}$, so by Proposition 67, $a^{(p-1)/2} \equiv \pm 1 \pmod{p}$. And since $\left(\frac{a}{p}\right) = \pm 1$, we have that $\left(\frac{a}{p}\right) \equiv \pm 1 \pmod{p}$. Let g be a primitive root modulo p . Then $a \equiv g^s \pmod{p}$ for some positive integer s . Hence

$$a^{(p-1)/2} \equiv 1 \pmod{p} \Leftrightarrow (g^s)^{(p-1)/2} \equiv 1 \pmod{p}$$

$$\Leftrightarrow (p-1) \mid s \left(\frac{p-1}{2}\right) \Leftrightarrow 2 \mid s \Leftrightarrow \left(\frac{a}{p}\right) = 1$$

by Proposition 90. And since p is odd, $\left(\frac{a}{p}\right) = 1$ if and only if $\left(\frac{a}{p}\right) \equiv 1 \pmod{p}$. \square

Proposition 92. *If a and b are integers and p is a prime, then*

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right) \left(\frac{b}{p}\right).$$

Proof. Let us first consider the case where $p \mid ab$. Then $\left(\frac{ab}{p}\right) = 0$ by definition. And by Proposition 29 either $p \mid a$ (in which case $\left(\frac{a}{p}\right) = 0$) or $p \mid b$ (in which case $\left(\frac{b}{p}\right) = 0$); thus $\left(\frac{a}{p}\right)\left(\frac{b}{p}\right) = 0$.

Suppose from now on that $p \nmid ab$. Then $p \nmid a$ and $p \nmid b$. By Proposition 91,

$$\left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \pmod{p} \quad \text{and} \quad \left(\frac{b}{p}\right) \equiv b^{(p-1)/2} \pmod{p}.$$

Then

$$\left(\frac{a}{p}\right)\left(\frac{b}{p}\right) \equiv a^{(p-1)/2} \cdot b^{(p-1)/2} \equiv (ab)^{(p-1)/2} \pmod{p},$$

and this last is congruent modulo p to $\left(\frac{ab}{p}\right)$ by Proposition 91. Since p is odd (and therefore $p \nmid (1 - (-1)) = 2$), it follows that $\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$. □

Now that we have proven some of the basic properties of the Legendre symbol, let us turn our attention to finding an efficient way to compute it. The most powerful tool for doing this is known as Gauss' law of quadratic reciprocity. The following two propositions are useful in the proof of this law.

Proposition 93. (*Lemma of Gauss*) Let p be an odd prime, a an integer such that $p \nmid a$. Consider the least residues of

$$1 \cdot a, 2 \cdot a, \dots, \left(\frac{p-1}{2}\right) \cdot a$$

modulo p . If exactly n of these least residues exceed $p/2$, then $\left(\frac{a}{p}\right) = (-1)^n$.

Proposition 94. If p is an odd prime, a is an odd integer such that $p \nmid a$, and we define t by

$$t = \sum_{j=1}^{\frac{p-1}{2}} \left\lfloor \frac{ja}{p} \right\rfloor,$$

then $\left(\frac{a}{p}\right) = (-1)^t$. Moreover, $\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}$.

Theorem 95. (Gauss' law of quadratic reciprocity) If p and q are distinct odd primes, then

$$\left(\frac{p}{q}\right) \left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2} \frac{q-1}{2}}.$$

Proof. Define

$$S = \{(x, y) \mid x, y \in \mathbb{Z}, 1 \leq x \leq \frac{p-1}{2}, 1 \leq y \leq \frac{q-1}{2}\},$$

$$S_1 = \{(x, y) \in S \mid qx > py\},$$

$$S_2 = \{(x, y) \in S \mid qx < py\}.$$

Notice that if $(x, y) \in S$, then $qx \neq py$, for if $qx = py$, then $p \mid qx$, so by Proposition 29, $p \mid q$ or $p \mid x$. But p and q are distinct primes, so $p \nmid q$, and since $1 \leq x \leq \frac{p-1}{2}$, $p \nmid x$, a contradiction. It follows that $S_1 \cup S_2 = S$. Moreover, it is clear that $S_1 \cap S_2 = \emptyset$. Therefore $\#S = \#S_1 + \#S_2$.

Since

$$\#S_1 = \sum_{j=1}^{\frac{p-1}{2}} \left\lfloor \frac{qj}{p} \right\rfloor \quad \text{and} \quad \#S_2 = \sum_{j=1}^{\frac{q-1}{2}} \left\lfloor \frac{pj}{q} \right\rfloor,$$

we have by Proposition 94 that

$$\left(\frac{p}{q}\right) \left(\frac{q}{p}\right) = (-1)^{\#S_1 + \#S_2} = (-1)^{\#S} = (-1)^{\frac{p-1}{2} \frac{q-1}{2}}.$$

□

Corollary 96. Suppose p and q are distinct odd primes.

- (1) If $p \equiv 1 \pmod{4}$ or $q \equiv 1 \pmod{4}$, then $\left(\frac{p}{q}\right) = \left(\frac{q}{p}\right)$.
- (2) If $p \equiv 3 \pmod{4}$ and $q \equiv 3 \pmod{4}$, then $\left(\frac{p}{q}\right) = -\left(\frac{q}{p}\right)$.

In the next proposition we collect results that, along with Theorem 95 and Proposition 92, allow us to compute Legendre symbols with great efficiency.

Proposition 97. *Let p be an odd prime.*

$$(1) \left(\frac{-1}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1 \pmod{4} \\ -1 & \text{if } p \equiv 3 \pmod{4} \end{cases}$$

$$(2) \left(\frac{2}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1, 7 \pmod{8} \\ -1 & \text{if } p \equiv 3, 5 \pmod{8} \end{cases}$$

Proof. We already proved the first part above. For the second part, since p is odd, $p \equiv 1, 3, 5$ or $7 \pmod{8}$. If $p \equiv 1 \pmod{8}$, then $p = 8k + 1$ for some positive integer k . Then by Proposition 94,

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}} = (-1)^{8k^2+2k} = 1.$$

If $p \equiv 3 \pmod{8}$, then $p = 8k + 3$ and

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}} = (-1)^{8k^2+6k+1} = -1.$$

If $p \equiv 5 \pmod{8}$, then $p = 8k + 5$ and

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}} = (-1)^{8k^2+10k+3} = -1.$$

Finally, if $p \equiv 7 \pmod{8}$, then $p = 8k + 7$ and

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}} = (-1)^{8k^2+14k+6} = 1.$$

□

Example 98. To illustrate the use quadratic reciprocity, let us compute $\left(\frac{-70}{103}\right)$. By Proposition 92,

$$\left(\frac{-70}{103}\right) = \left(\frac{-1}{103}\right) \left(\frac{2}{103}\right) \left(\frac{5}{103}\right) \left(\frac{7}{103}\right).$$

Since $103 \equiv 3 \pmod{4}$, Proposition 97 (1) implies that $\left(\frac{-1}{103}\right) = -1$, and since $103 \equiv 7 \pmod{8}$, Proposition 97 (2) implies that $\left(\frac{2}{103}\right) = 1$. Moreover, Corollary 96 implies that

$$\left(\frac{5}{103}\right) = \left(\frac{103}{5}\right) = \left(\frac{3}{5}\right) = -1$$

and

$$\left(\frac{7}{103}\right) = -\left(\frac{103}{7}\right) = -\left(\frac{5}{7}\right) = -\left(\frac{7}{5}\right) = -\left(\frac{2}{5}\right) = 1.$$

Thus $\left(\frac{-70}{103}\right) = (-1)(1)(-1)(1) = 1$.

Example 99. We can use quadratic reciprocity to completely characterize all the primes p such that $\left(\frac{3}{p}\right) = 1$. First note that $\left(\frac{3}{2}\right) = 1$ and $\left(\frac{3}{3}\right) = 0$. Since $3 \equiv 3 \pmod{4}$, Corollary 96 tells us that $\left(\frac{3}{p}\right) = \left(\frac{p}{3}\right)$ if $p \equiv 1 \pmod{4}$ and $\left(\frac{3}{p}\right) = -\left(\frac{p}{3}\right)$ if $p \equiv 3 \pmod{4}$. Moreover, one can easily check that 1 is a quadratic residue modulo 3 and that 2 is a nonresidue modulo 3. Hence $\left(\frac{p}{3}\right) = 1$ if $p \equiv 1 \pmod{3}$ and $\left(\frac{p}{3}\right) = -1$ if $p \equiv 2 \pmod{3}$. Putting all of this together using the Chinese Remainder Theorem, we find that if p is prime, then

$$\left(\frac{3}{p}\right) = \begin{cases} 0 & \text{if } p = 3, \\ 1 & \text{if } p \equiv 1 \text{ or } 11 \pmod{12}, \text{ and} \\ -1 & \text{if } p \equiv 5 \text{ or } 7 \pmod{12}. \end{cases}$$

DESCENT

The method of descent, which was first practiced by Fermat, is based on the *well-ordering principle*, which states that every nonempty set of positive integers has a least element. Indeed, the method of descent uses explicitly the converse of this: if S is a set of positive integers such that for every $a \in S$ there is a $b \in S$ with $b < a$, then $S = \emptyset$.

We illustrate the method of descent by giving a proof of

Theorem 100. (*Fermat*) *Given an odd prime p , there exist integers x and y such that $p = x^2 + y^2$ if and only if $p \equiv 1 \pmod{4}$.*

Our proof of Theorem 100 is based on the following proposition.

Proposition 101. *Suppose N is the sum of relatively prime squares (i.e., there exist $a, b \in \mathbb{Z}$ such that $(a, b) = 1$ and $N = a^2 + b^2$), and suppose that q is a prime such that $q \mid N$ and q is the sum of relatively prime squares. Then N/q is the sum of relatively prime squares.*

Proof. Write $q = x^2 + y^2$. Then by Proposition 7 (2),

$$\begin{aligned} q \mid (x^2N - a^2q) &= x^2(a^2 + b^2) - a^2(x^2 + y^2) \\ &= x^2b^2 - a^2y^2 = (xb - ay)(xb + ay). \end{aligned}$$

By Proposition 29, $q \mid (xb - ay)$ or $q \mid (xb + ay)$. Replacing a by $-a$ if necessary, we may assume that $q \mid (xb - ay)$. So $xb - ay = dq$ for some $d \in \mathbb{Z}$.

We claim that $x \mid (a + dy)$. To see this, note that

$$\begin{aligned} (a + dy)y &= ay + dy^2 = xb - dq + dy^2 = xb - d(x^2 + y^2) + dy^2 \\ &= xb - dx^2 = x(b - dx), \end{aligned}$$

so by definition $x \mid (a + dy)y$. By assumption $(x, y) = 1$, and hence by Proposition 49 (3), $x \mid (a + dy)$. Write $a + dy = cx$; note that by our work above, $b = dx + cy$. Then

$$\begin{aligned} N = a^2 + b^2 &= (cx - dy)^2 + (dx + cy)^2 \\ &= c^2x^2 - 2cxdy + d^2y^2 + d^2x^2 + 2dxcy + c^2y^2 \\ &= c^2x^2 + d^2x^2 + d^2y^2 + c^2y^2 = x^2(c^2 + d^2) + y^2(d^2 + c^2) \end{aligned}$$

$$= (x^2 + y^2)(c^2 + d^2) = q(c^2 + d^2),$$

i.e., $N/q = c^2 + d^2$.

Finally, recall that $a = cx - dy$ and $b = dx + cy$. By Proposition 7 (2), $(c, d) \mid a$ and $(c, d) \mid b$. Therefore $(c, d) \leq (a, b) = 1$, and thus $(c, d) = 1$. \square

Now let us prove Theorem 100. First suppose that $p = x^2 + y^2$. Since p is odd, $p \equiv 1$ or $3 \pmod{4}$. And because each of x^2 and y^2 is $\equiv 0$ or $1 \pmod{4}$, it cannot be the case that $x^2 + y^2 = p \equiv 3 \pmod{4}$. Thus $p \equiv 1 \pmod{4}$.

Let us prove the reverse implication by contradiction. Let S denote the set of primes that are congruent to 1 modulo 4 and are not the sum of squares, and assume that S is nonempty. Then S has a least element; denote it by p . By Proposition 70, there exists an integer x such that $x^2 \equiv -1 \pmod{p}$; of course, $-x$ is also a solution. Denote by b the least residue of x modulo p . Since p is odd, the set $\{b, p - b\}$ consists of one odd number and one even number; let a be the even element of this set. Since $a \equiv \pm b \equiv \pm x \pmod{p}$, we have that $a^2 \equiv -1 \pmod{p}$.

Let $N = a^2 + 1$. Then $p \mid N$, and since $N = a^2 + 1^2$, N is a sum of relatively prime squares. Because N is odd, by Proposition 70, every prime divisor q of N is congruent to 1 modulo 4 (for if $q \mid N = a^2 + 1$, then q is odd and $a^2 \equiv -1 \pmod{q}$). Moreover, since $a < p$, $N \leq (p - 1)^2 + 1 < p^2$. It follows that if q is a prime divisor of N and $q \neq p$, then $q < p$. By the minimality of p , each such $q \neq p$ can be written as a sum of squares (note that since q is prime, these will be relatively prime squares). Therefore, by Proposition 101, N/q can be written as a sum of relatively prime squares. Repeating this argument on N/q , etc., we can eliminate all prime divisors of N except for p and conclude that p is a sum of relatively prime squares, a contradiction.

QUADRATIC FORMS

The polynomial $x^2 + y^2$ that we studied in the last section is an example of what we call a binary (because it involves two variables) quadratic (because every term has degree two) form. The study of quadratic forms was one of the central subjects in the early study of number theory.

Definition 102. A *binary quadratic form* is a polynomial of the form

$$f(x, y) = ax^2 + bxy + cy^2$$

with $a, b, c \in \mathbb{Z}$. The *discriminant* of f is $d = b^2 - 4ac$.

Assume now that $a \neq 0$. If we complete the square, we obtain

$$\begin{aligned} f(x, y) &= a \left(x^2 + \frac{b}{a}xy \right) + cy^2 \\ &= a \left(x^2 + \frac{b}{a}xy + \frac{b^2}{4a^2}y^2 \right) + cy^2 - \left(\frac{b^2}{4a} \right) y^2 \\ &= a \left(x + \frac{b}{2a}y \right)^2 + \left(\frac{4ac - b^2}{4a} \right) y^2 \\ &= a \left(x + \frac{b}{2a}y \right)^2 - \left(\frac{d}{4a} \right) y^2 \end{aligned}$$

Thus if $a > 0$ and $d < 0$, then $f(x, y)$ takes only nonnegative values.

Definition 103. If $f(x, y) = ax^2 + bxy + cy^2$ is a quadratic form, we say that f is *positive definite* if $a > 0$ and $d < 0$.

Notice that if $d > 0$, then f will take both positive and negative values, regardless of the sign of a .

Proposition 104. *Suppose d is an integer. Then there exists a binary quadratic form with discriminant d if and only if $d \equiv 0$ or $1 \pmod{4}$.*

Proof. If d is a discriminant, then

$$d = b^2 - 4ac \equiv b^2 \equiv 0 \text{ or } 1 \pmod{4}.$$

Now suppose $d \equiv 0$ or $1 \pmod{4}$. If $d \equiv 0 \pmod{4}$, then the form $x^2 - (d/4)y^2$ has discriminant d , and if $d \equiv 1 \pmod{4}$, then the form $x^2 + xy - \left(\frac{d-1}{4}\right)y^2$ has discriminant d . \square

We discovered in the last section exactly which primes are of the form $x^2 + y^2$. Let us now consider the general question of the prime values taken by quadratic forms.

Definition 105. We say that the quadratic form $f(x, y)$ represents an integer n if there exist integers x_0 and y_0 such that $f(x_0, y_0) = n$. Such a representation is called *proper* if $(x_0, y_0) = 1$ and *improper* if $(x_0, y_0) > 1$.

Theorem 106. Let $n, d \in \mathbb{Z}$, and suppose that $n \neq 0$. Then there exists a binary quadratic form of discriminant d that properly represents n if and only if there exists an integer b such that $b^2 \equiv d \pmod{4|n|}$.

Proof. First suppose that there exists an integer b such that $b^2 \equiv d \pmod{4|n|}$. Then $b^2 - d = 4nk$ for some integer k . Hence $f(x, y) = nx^2 + bxy + ky^2$ is a binary quadratic form of discriminant $b^2 - 4nk = d$; moreover, $f(1, 0) = n$, and clearly this is a proper representation.

Now suppose that there exists a binary quadratic form $f(x, y) = ax^2 + bxy + cy^2$ of discriminant d that properly represents n . So there exist integers x_0 and y_0 such that $f(x_0, y_0) = n$ and $(x_0, y_0) = 1$. Let

$$m_1 = \prod_{p|(4|n|), p|x_0} p^{v_p(4|n|)}$$

and $m_2 = (4|n|)/m_1$. Clearly $4|n| = m_1 m_2$ and $(m_1, m_2) = 1$, and since $(x_0, y_0) = 1$, it follows that $(m_1, y_0) = (m_2, x_0) = 1$. If we complete the square on f with respect to x as we did above:

$$f(x, y) = a \left(x + \frac{b}{2a} y \right)^2 - \left(\frac{d}{4a} \right) y^2,$$

we find that

$$4an = 4af(x_0, y_0) = (2ax_0 + by_0)^2 - dy_0^2,$$

and therefore $(2ax_0 + by_0)^2 \equiv dy_0^2 \pmod{m_1}$. Since $(y_0, m_1) = 1$, by Proposition 42 there exists an integer r such that $ry_0 \equiv 1 \pmod{m_1}$. Thus

$$\begin{aligned} [(2ax_0 + by_0)r]^2 &\equiv r^2(2ax_0 + by_0)^2 \\ &\equiv r^2 dy_0^2 \equiv (ry_0)^2 d \equiv d \pmod{m_1}. \end{aligned}$$

We see that the congruence $x^2 \equiv d \pmod{m_1}$ has a solution.

On the other hand, if we complete the square on f with respect to y instead of x , we obtain

$$f(x, y) = c \left(y + \frac{b}{2c} x \right)^2 - \left(\frac{d}{4c} \right) x^2,$$

which implies that

$$4cn = 4cf(x_0, y_0) = (2cy_0 + bx_0)^2 - dx_0^2,$$

and therefore $(2cy_0 + bx_0)^2 \equiv dx_0^2 \pmod{m_2}$. Since $(x_0, m_2) = 1$, by Proposition 42 there exists an integer s such that $sx_0 \equiv 1 \pmod{m_2}$. Thus

$$\begin{aligned} [(2cy_0 + bx_0)s]^2 &\equiv s^2(2cy_0 + bx_0)^2 \\ &\equiv s^2dx_0^2 \equiv (sx_0)^2d \equiv d \pmod{m_2}. \end{aligned}$$

We see that the congruence $x^2 \equiv d \pmod{m_2}$ has a solution; recall that we already showed that the congruence $x^2 \equiv d \pmod{m_1}$ has a solution. Since $(m_1, m_2) = 1$ and $m_1m_2 = 4|n|$, by the Chinese Remainder Theorem we know that the congruence $x^2 \equiv d \pmod{4|n|}$ has a solution, as required. \square

Now we can give a criterion for when a prime is represented by a form of a given discriminant.

Corollary 107. *Suppose that d is an integer such that $d \equiv 0$ or $1 \pmod{4}$, and that p is an odd prime. Then there is a binary quadratic form of discriminant d that represents p if and only if $p \mid d$ or $\left(\frac{d}{p}\right) = 1$.*

Proof. First let us note that a representation of p by a quadratic form $f(x, y)$ is proper, because if $f(x_0, y_0) = p$, then $(x_0, y_0)^2 \mid f(x_0, y_0) = p$, and therefore $(x_0, y_0) = 1$.

Suppose that p is represented by a binary quadratic form of discriminant d . Since it is therefore properly represented, by Theorem 106 the congruence $b^2 \equiv d \pmod{4p}$ has a solution. If b is a solution, then $4p \mid (b^2 - d)$, and therefore $p \mid (b^2 - d)$, i.e., $b^2 \equiv d \pmod{p}$. It follows that either $p \mid d$ or $\left(\frac{d}{p}\right) = 1$.

Now suppose that $p \mid d$ or $\left(\frac{d}{p}\right) = 1$. Then there exists an integer b such that $b^2 \equiv d \pmod{p}$ (note that if $p \mid d$, then $b \equiv 0 \pmod{p}$). And since $d \equiv 0$ or $1 \pmod{4}$, either $0^2 \equiv d \pmod{4}$ or $1^2 \equiv d \pmod{4}$, i.e. the congruence $x^2 \equiv d \pmod{4}$ has a solution. Since p is odd, $(p, 4) = 1$, so by the Chinese Remainder Theorem the congruence $b^2 \equiv d \pmod{4p}$ has a solution. By Theorem 106, then, there exists a binary quadratic form of discriminant d that properly represents p . \square

The only weakness of Corollary 107 is that given a discriminant d , one can construct infinitely many triples of integers (a, b, c) such that $f(x, y) = ax^2 + bxy + cy^2$ has discriminant d .

For example, say that we are interested in the question of which primes are represented by the form $x^2 + 3y^2$. Since $x^2 + 3y^2 \equiv x^2 \equiv 0$ or $1 \pmod{3}$, the only primes that can be represented by $x^2 + 3y^2$ are 3 and those primes that are congruent to 1 modulo 3. A quick check shows that the first several primes congruent to 1 modulo 3 are represented by $f(x, y) = x^2 + 3y^2$:

$$7 = f(2, 1), \quad 13 = f(1, 2), \quad 19 = f(4, 1), \quad 31 = f(2, 3).$$

A natural question is “are ALL primes congruent to 1 modulo 3 represented by $x^2 + 3y^2$?” Well, the discriminant of $x^2 + 3y^2$ is -12 , and if $p \equiv 1 \pmod{3}$, Corollary 107 tells us that p is represented by a form of discriminant -12 . To see this, note that

$$\left(\frac{-12}{p}\right) = \left(\frac{-3}{p}\right) \left(\frac{4}{p}\right) = \left(\frac{-3}{p}\right) = \left(\frac{-1}{p}\right) \left(\frac{3}{p}\right).$$

Now, by Corollary 96, $\left(\frac{3}{p}\right) = \left(\frac{p}{3}\right)$ if $p \equiv 1 \pmod{4}$ and $\left(\frac{3}{p}\right) = -\left(\frac{p}{3}\right)$ if $p \equiv 3 \pmod{4}$. And by Proposition 97 (1), $\left(\frac{-1}{p}\right) = 1$ if $p \equiv 1 \pmod{4}$ and $\left(\frac{-1}{p}\right) = -1$ if $p \equiv 3 \pmod{4}$. In either case, then, $\left(\frac{-1}{p}\right) \left(\frac{3}{p}\right) = \left(\frac{p}{3}\right) = \left(\frac{1}{3}\right) = 1$.

As we mentioned, the problem is that one can easily write down lots of quadratic forms that have discriminant -12 , for example

$$x^2 + 2xy + 4y^2, \quad 7x^2 - 4xy + y^2, \quad \dots$$

Unless we deal with this ambiguity, then, it will be the case that for all we know, $x^2 + 3y^2$ represents only some of the primes $p \equiv 1 \pmod{3}$, while $x^2 + 2xy + 4y^2$ and $7x^2 - 4xy + y^2$ represent the rest.

We resolve the ambiguity as follows.

Definition 108. We say that the binary quadratic forms $f(x, y)$ and $g(x, y)$ are *equivalent*, and write $f \sim g$, if there exist integers α, β, γ and δ such that $\alpha\delta - \beta\gamma = 1$ and $f(\alpha x + \beta y, \gamma x + \delta y) = g(x, y)$. In this case we say that $(\alpha, \beta, \gamma, \delta)$ takes f to g .

Example 109. $x^2 + 3y^2 \sim x^2 + 2xy + 4y^2$ because $(1, 1, 0, 1)$ takes $x^2 + 3y^2$ to $(x + y)^2 + 3(y)^2 = x^2 + 2xy + 4y^2$.

Proposition 110. *Suppose f and g are binary quadratic forms, and that $f \sim g$. Then f and g have the same discriminant.*

Proof. Write $f(x, y) = ax^2 + bxy + cy^2$, and suppose $(\alpha, \beta, \gamma, \delta)$ takes f to g . Then f has discriminant $b^2 - 4ac$, and

$$g(x, y) = f(\alpha x + \beta y, \gamma x + \delta y)$$

$$= (a\alpha^2 + b\alpha\gamma + c\gamma^2)x^2 + (2a\alpha\beta + b\alpha\gamma + b\beta\gamma + 2c\gamma\delta)xy + (a\beta^2 + b\beta\delta + c\delta^2)y^2.$$

It is left as a straightforward, but fairly painful, exercise to show that this form has discriminant $b^2 - 4ac$. □

Proposition 111. *Let f , g and h be binary quadratic forms. Then*

- (1) $f \sim f$,
- (2) if $f \sim g$, then $g \sim f$, and
- (3) if $f \sim g$ and $g \sim h$, then $f \sim h$.

Proof. For (1), note that $(1, 0, 0, 1)$ takes f to f .

For (2), suppose $(\alpha, \beta, \gamma, \delta)$ takes f to g . Then $f(\alpha x + \beta y, \gamma x + \delta y) = g(x, y)$. We claim that $(\delta, -\beta, -\gamma, \alpha)$ (notice that $\delta\alpha - (-\beta)(-\gamma) = 1$) takes g to f ; to see this, note that

$$g(\delta x - \beta y, -\gamma x + \alpha y) =$$

$$f(\alpha(\delta x - \beta y) + \beta(-\gamma x + \alpha y), \gamma(\delta x - \beta y) + \delta(-\gamma x + \alpha y))$$

$$= f(x, y),$$

since $\alpha\delta - \beta\gamma = 1$.

Finally, for (3), if $(\alpha, \beta, \gamma, \delta)$ takes f to g and $(\alpha', \beta', \gamma', \delta')$ takes g to h , then

$$h(x, y) = g(\alpha'x + \beta'y, \gamma'x + \delta'y)$$

$$= f(\alpha(\alpha'x + \beta'y) + \beta(\gamma'x + \delta'y), \gamma(\alpha'x + \beta'y) + \delta(\gamma'x + \delta'y))$$

$$= f((\alpha\alpha' + \beta\gamma')x + (\alpha\beta' + \beta\delta')y, (\gamma\alpha' + \delta\gamma')x + (\gamma\beta' + \delta\delta')y),$$

and so $(\alpha\alpha' + \beta\gamma', \alpha\beta' + \beta\delta', \gamma\alpha' + \delta\gamma', \gamma\beta' + \delta\delta')$ takes f to h . □

From now on we will consider two binary quadratic forms the same if they are equivalent in the above sense (i.e., we will consider equivalence classes of forms). Since we are concerned with which numbers forms represent, this is justified by the following result.

Proposition 112. *Suppose that f and g are equivalent binary quadratic forms. Then f and g represent exactly the same integers. Moreover, if f represents an integer n , then the representations of n by f are in one-to-one correspondence with the representations of n by g . Even further over, the proper representations of n by f and g are in one-to-one correspondence.*

Proof. Suppose $(\alpha, \beta, \gamma, \delta)$ takes f to g . Recall that $(\delta, -\beta, -\gamma, \alpha)$ takes g to f . So if $g(x_0, y_0) = n$, then $f(x'_0, y'_0) = n$, where $x'_0 = \alpha x_0 + \beta y_0$ and $y'_0 = \gamma x_0 + \delta y_0$. Notice that we can recover our representation of n by g using $(\delta, -\beta, -\gamma, \alpha)$, since

$$\delta x'_0 - \beta y'_0 = (\alpha\delta - \beta\gamma)x_0 + (\delta\beta - \beta\delta)y_0 = x_0$$

and

$$-\gamma x'_0 + \alpha y'_0 = (-\gamma\alpha + \alpha\gamma)x_0 + (-\gamma\beta + \alpha\delta)y_0 = y_0.$$

Thus the representations of n by f and g are in one-to-one correspondence.

Finally, for the proper part, we just need to show that our correspondence takes proper representations to proper representations. To see this, suppose that $(x_0, y_0) = 1$, and set $d = (\alpha x_0 + \beta y_0, \gamma x_0 + \delta y_0)$. We claim that $d = 1$. To see this, we simply note that by Proposition 7 (2),

$$d \mid (\delta(\alpha x_0 + \beta y_0) - \beta(\gamma x_0 + \delta y_0)) = x_0$$

and

$$d \mid (-\gamma(\alpha x_0 + \beta y_0) + \alpha(\gamma x_0 + \delta y_0)) = y_0.$$

It follows that $d \leq (x_0, y_0) = 1$, and therefore $d = 1$. □

Example 113. Recall that we saw above that the first several primes $p \equiv 1 \pmod{3}$ are represented by $f(x, y) = x^2 + 3y^2$:

$$7 = f(2, 1), \quad 13 = f(1, 2), \quad 19 = f(4, 1), \quad 31 = f(2, 3).$$

And we also saw that f is equivalent to $g(x, y) = x^2 + 2xy + 4y^2$, and that $(1, 1, 0, 1)$ takes f to g (i.e., $f(x + y, y) = g(x, y)$). Using this

transformation we may easily make all of these representations by f into representations by g :

$$7 = f(2, 1) = g(2-1, 1) = g(1, 1), \quad 13 = f(1, 2) = g(1-2, 2) = g(-1, 2),$$

$$19 = f(4, 1) = g(4-1, 1) = g(3, 1), \quad 31 = f(2, 3) = g(2-3, 3) = g(-1, 3).$$

Since for the purposes of representation equivalent forms are the same by Proposition 112, we may as well choose an element from each equivalence class to stand for the entire class.

Definition 114. Let $f(x, y) = ax^2 + bxy + cy^2$ be a binary quadratic form whose discriminant $d = b^2 - 4ac$ is not a perfect square. We say that f is *reduced* if

$$-|a| < b \leq |a| < |c|$$

or

$$0 \leq b \leq |a| = |c|.$$

Proposition 115. *Suppose d is an integer such that $d \equiv 0$ or $1 \pmod{4}$ and d is not a perfect square. Let $f(x, y) = ax^2 + bxy + cy^2$ be a binary quadratic form of discriminant d . Then f is equivalent to a reduced form.*

Proof. Since $d = b^2 - 4ac$ is not a square, $a, c \neq 0$ (else $d = b^2$). We now describe a procedure that will produce a reduced form equivalent to f .

Notice first that $(1, m, 0, 1)$ takes f to

$$a(x + my)^2 + b(x + my)y + cy^2 = ax^2 + (b + 2am)xy + (am^2 + bm + c)y^2.$$

Therefore there is a unique m such that $-|a| < b + 2am \leq |a|$. Upon using this transformation with this unique m , then, we may assume that $-|a| < b \leq |a|$.

If $|a| < |c|$, our form is reduced and we stop. If $|a| = |c|$ and $0 \leq b \leq |a|$, our form is reduced and we stop. If neither of these is the case, let us apply $(0, 1, -1, 0)$. This takes f to

$$a(y)^2 + b(y)(-x) + c(-x)^2 = cx^2 - bxy + ay^2.$$

If it is the case that $|a| = |c|$ and $-|a| < b < 0$, this is a reduced form and we stop. If $|a| > |c|$, we repeat this procedure.

Notice that each application of this procedure either produced a reduced form OR reduces the size of the coefficient of x^2 . It follows that we will obtain a reduced form after a finite number of applications of this procedure. □

Proposition 116. *Let $f(x, y) = ax^2 + bxy + cy^2$ be a reduced positive definite binary quadratic form (i.e., $d = b^2 - 4ac < 0$ and $a, c > 0$). Then $a \leq \sqrt{-d/3}$.*

Proof. Since $|b| \leq |a| \leq |c|$, we have that

$$d = b^2 - 4ac \leq a^2 - 4ac \leq a^2 - 4a^2 = -3a^2,$$

which implies that $3a^2 \leq -d$, from which our result follows. □

Corollary 117. *Given a discriminant d which is not a perfect square, there are finitely many reduced forms of discriminant d .*

Proof. If $f(x, y) = ax^2 + bxy + cy^2$ is a reduced form of discriminant d , then by Proposition 116 there are only finitely many values that a can take. Since $|b| \leq |a|$, there are also only finitely many values that b can take. And once a and b are given, c is determined by $d = b^2 - 4ac$. Thus there are finitely many possible triples (a, b, c) which can produce a reduced form of discriminant d . □

Example 118. Let us find all of the reduced positive definite binary quadratic forms of discriminant -12 . As we are considering positive definite forms, $a > 0$, and by Proposition 116, $a \leq \sqrt{4} = 2$. Hence there are only two possible values for a : 1 and 2. If $a = 1$, then since we are looking for reduced forms, $|b| \leq |a| = 1$, so $b = -1, 0$ or 1 . Substituting these values in $-12 = b^2 - 4ac$, the only viable b value is 0, which gives $c = 3$. This produces the reduced form $x^2 + 3y^2$. If $a = 2$, then $|b| \leq |a| = 2$, so $b = -2, -1, 0, 1$ or 2 . Substituting these into $-12 = b^2 - 4ac$, we find that the $b = \pm 2$ values are viable, giving a c value of 2. This produces the forms $2x^2 \pm 2xy + 2y^2$, of which only the form with positive middle coefficient is reduced. Thus we find that the reduced forms of discriminant -12 are $x^2 + 3y^2$ and $2x^2 + 2xy + 2y^2$.

The final important property of reduced forms is the following.

Proposition 119. *Suppose that f and g are reduced positive definite binary quadratic forms. If $f \sim g$, then $f = g$.*

Definition 120. If $d \equiv 0$ or $1 \pmod{4}$ and $d < 0$, we denote by $H(d)$ the number of reduced positive definite binary quadratic forms of discriminant d .

Notice that Proposition 119 tells us that $H(d)$ is also equal to the number of equivalence classes of reduced positive definite binary quadratic forms of discriminant d .

Example 121. By Example 118, $H(-12) = 2$.

ELLIPTIC CURVES

The subject of elliptic curves is one of the most active areas of current mathematical research. The theory of elliptic curves turns up in diverse and perhaps surprising places: cryptography, factoring, and the proof of Fermat's Last Theorem, to name a few.

Definition 122. An elliptic curve E defined over \mathbb{Q} is given by an equation of the form

$$E \quad : \quad y^2 = x^3 + ax^2 + bx + c = f(x),$$

where $a, b, c \in \mathbb{Z}$ and $f(x)$ and $f'(x)$ have no common root.

Example 123. The equation $y^2 = x^3 - 3x$ defines an elliptic curve, since the roots of $f'(x) = 3x^2 - 3$ are ± 1 , and $f(1) = -2 \neq 0$, $f(-1) = 2 \neq 0$.

Nonexample 124. The equation $y^2 = x^3 - 2x^2 + x$ does not define an elliptic curve, since $f'(x) = 3x^2 - 4x + 1$ has $x = 1$ as a root, and $f(1) = 0$.

Nonexample 125. The equation $y^2 = x^3$ does not define an elliptic curve, since $f'(x) = 3x^2$ has $x = 0$ as a root, and $f(0) = 0$.

When an elliptic curve is drawn in the real plane, it has either one or two components, depending upon whether f has one or three real roots (note that f will not have exactly two real roots, for if it did, it would intersect the x -axis tangentially at a point, and the x -coordinate of that point would be a root of both f and f').

When we draw the graph of a relation in the plane, what we are drawing is sometimes called the "affine" part of the graph. Elliptic curves as we have defined them are also endowed with a "point at infinity", which we will denote by ∞ . We include this point because elliptic curves are really "projective curves", not affine curves. It is not the point of this course to delve into projective geometry; suffice it to say that the projective plane contains the affine plane, and that as we have defined them, elliptic curves consist of their affine part plus one point of the projective plane that is not part of the affine plane.

The key property of ∞ is the following: if we take a line in the affine plane and consider it projectively, it attains one more point (just as E does), and this point is ∞ if and only if the line is vertical.

If E is an elliptic curve defined over \mathbb{Q} as above, we define

$$E(\mathbb{Q}) = \{(x, y) \mid x, y \in \mathbb{Q}, \quad y^2 = x^3 + ax^2 + bx + c\} \cup \{\infty\}.$$

We call $E(\mathbb{Q})$ the *rational points of E* .

Recall that we were able to parametrize the rational points on the unit circle by fixing a rational point (we chose $(-1, 0)$, but any rational point would do), taking all lines through that point having rational slope, and finding the second point of intersection of those lines with the unit circle. Elliptic curves CANNOT be parametrized in the same way, and indeed, cannot be parametrized at all. To get an idea of why this is so, consider for a moment the elliptic curve $E : y^2 = x^3 - x$. Notice that this curve has $(-1, 0)$ as a rational point. Say we use the same idea of taking all lines through this point having rational slope. For example, if we take slope 1, we get the line $y = x + 1$. Substituting, we obtain

$$(x + 1)^2 = x^3 - x \iff 0 = x^3 - x^2 - 3x - 1 = (x + 1)(x^2 - 2x - 1).$$

We see that this line intersects E in two other points, whose x -coordinates are $1 \pm \sqrt{2}$, the roots of the quadratic. Therefore the other two points of intersection of this line with E are NOT rational points.

A natural question at this point is: is there any way that, given a rational point (or points) on an elliptic curve E , we can produce more rational points on E ? The answer to this question is “yes”, but while with the unit circle we required but one rational point to produce ALL the others (the algebraic reason for this being that the unit circle is defined by a quadratic equation), on an elliptic curve we require two rational points (since an elliptic curve is defined by a cubic equation) merely to produce ONE MORE rational point.

Proposition 126. *Suppose that the polynomial $(x - \alpha)(x - \beta)(x - \gamma)$ has rational coefficients when it is multiplied out. If $\alpha, \beta \in \mathbb{Q}$, then $\gamma \in \mathbb{Q}$.*

Proof. If we multiply out the polynomial, the x^2 coefficient (call it δ) is equal to $-(\alpha + \beta + \gamma)$. Since $\alpha, \beta, \delta \in \mathbb{Q}$, it follows that $\gamma = \delta + \alpha + \beta \in \mathbb{Q}$. \square

Suppose $P = (x_P, y_P)$, $Q = (x_Q, y_Q)$ are distinct rational points on E with $x_P \neq x_Q$. The slope of the line through P and Q is rational (as it is equal to $(y_Q - y_P)/(x_Q - x_P)$), and hence if we denote the equation of the line through P and Q by $y = \alpha x + \beta$, we have that $\alpha, \beta \in \mathbb{Q}$. To find the third point of intersection of this line with E , we substitute

$$(\alpha x + \beta)^2 = x^3 + ax^2 + bx + c,$$

and if we move all terms to the right hand side we see that we will have a cubic polynomial having rational coefficients set equal to zero.

By construction, two of the roots of this cubic are x_P and x_Q , which by assumption are rational. By Proposition 126, the third root is also rational, and substituting this into the equation $y = \alpha x + \beta$ gives a rational y , and thus the third point of intersection of the line with E is a rational point.

Example 127. Let $E : y^2 = x^3 + 17$. Then $(-2, 3)$ and $(4, 9)$ are rational points of E . Let us find the third point of intersection with E of the line through these two points.

The equation of this line is easily found to be $y = x + 5$, and substituting we obtain

$$(x + 5)^2 = x^3 + 17 \implies 0 = x^3 - x^2 - 10x - 8 = (x + 2)(x - 4)(x + 1).$$

We see that the x -coordinate of the third point of intersection is -1 , and substituting into $y = x + 5$ we find this third point to be $(-1, 4)$.

Example 128. Let us repeat the previous exercise with the same curve E , but with $P = (-2, 3)$ and $Q = (-2, 3)$. Since P and Q are the same here, there are infinitely many lines through P and Q ; which do we choose? Well, if P and Q were different, we'd take the unique line through them (sometimes called a secant line). Now imagine what happens to this secant line if we let Q approach P ; we know from calculus that in the limit, the secant line through P and Q becomes the tangent line to E at P , so this is the line that we take.

The slope of the tangent line is given by $y' = \frac{dy}{dx}$, so let us find this by implicit differentiation:

$$2yy' = 3x^2 \implies y' = \frac{3x^2}{2y}.$$

The slope of the tangent line to E at $(-2, 3)$ is therefore $3(-2)^2/2(3) = 2$, and so the equation of this line is $y = 2x + 7$. Substituting, we obtain

$$(2x + 7)^2 = x^3 + 17 \implies 0 = x^3 - 4x^2 - 28x - 32 = (x + 2)(x + 2)(x - 8).$$

We see that the x -coordinate of the third point of intersection is 8 , and substituting into $y = 2x + 7$ we find this third point to be $(8, 23)$.

Example 129. Let us continue with our theme by finding the third point of intersection with E of the line through $(-2, 3)$ and $(-2, -3)$. This line is vertical (with equation $x = -2$), and does not intersect E in any more affine points (since substituting $x = -2$ into the equation for E gives $y^2 = 9$), but it does intersect E in the point ∞ . Thus the third point of intersection is ∞ .

Example 130. Now let $P = (-2, 3)$ and $Q = \infty$. There are infinitely many lines through P , but as we have seen only one will contain ∞ : the vertical one. This line has equation $x = -2$, and substituting into the equation for E we obtain $y^2 = 9$, yielding $y = \pm 3$. Thus the third point of intersection is $(-2, -3)$.

Definition 131. Let E be an elliptic curve defined over \mathbb{Q} . If $P, Q \in E(\mathbb{Q})$ and (x, y) (respectively ∞) is the third point of intersection of the line through P and Q with E , we denote the point $(x, -y)$ (respectively ∞) by $P + Q$.

Note that $(x, -y)$ is simply the reflection of the point (x, y) across the x -axis.

Example 132. On $E : y^2 = x^3 + 17$, $(-2, 3) + (4, 9) = (-1, -4)$, $(-2, 3) + (-2, 3) = (8, -23)$, $(-2, 3) + (-2, -3) = \infty$, $(-2, 3) + \infty = (-2, 3)$

The notation $P + Q$ is certainly quite suggestive, and indeed this operation $+$ can be thought of as addition on $E(\mathbb{Q})$ in the sense that it has the same properties as addition in the integers: it is associative, commutative, has a neutral element, and every element can be negated. Indeed, looking at our examples above, we see that ∞ is the neutral element, and therefore the negative of (x, y) is $(x, -y)$. We will not prove associativity here in the interest of time and space (see the text, Lemma 5.20, for a proof if you wish). Commutativity is easy, since the line through P and Q is the same as the line through Q and P , and therefore by the definition of $+$ on $E(\mathbb{Q})$, $P + Q = Q + P$.

Since the operation $+$ on $E(\mathbb{Q})$ can truly be thought of as addition, we will denote $P + P = 2P$, $2P + P = 3P$, etc.

Example 133. On $E : y^2 = x^3 + 17$,

$$2(-2, 3) = (8, -23), \quad 3(-2, 3) = \left(\frac{19}{25}, \frac{522}{125} \right),$$

$$4(-2, 3) = \left(\frac{752}{529}, -\frac{54239}{12167} \right), \quad 5(-2, 3) = \left(\frac{174598}{32761}, \frac{76943337}{5929741} \right)$$

Indeed, if $m > n \geq 1$, then $m(-2, 3) \neq n(-2, 3)$ (we will not prove this). Therefore $E(\mathbb{Q})$ is infinite for this curve.

We saw above that $E(\mathbb{Q})$ is an infinite set for the curve $E : y^2 = x^3 + 17$. We note here that $E(\mathbb{Q})$ may be finite or infinite, depending upon the curve; for example, if $E : y^2 = x^3 - x$, it can be shown that $E(\mathbb{Q}) = \{\infty, (0, 0), (1, 0), (-1, 0)\}$.

Let us begin our study of elliptic curves “modulo p ”.

Definition 134. Let $E : y^2 = x^3 + ax^2 + bx + c = f(x)$ be an elliptic curve defined over \mathbb{Q} . Recall that this means that $a, b, c \in \mathbb{Z}$ and that $f(x)$ and $f'(x)$ have no common root. For any odd prime p we may consider $f(x)$ as a polynomial with coefficients in $\mathbb{Z}/p\mathbb{Z}$ (i.e., we may consider a, b and c as elements of $\mathbb{Z}/p\mathbb{Z}$). If it remains the case that $f(x)$ and $f'(x)$ have no common root, we say that E has *good reduction at p* ; otherwise we say that E has *bad reduction at p* .

Example 135. The elliptic curve $E : y^2 = x^3 - 2x$ has good reduction at $p = 5$. To see this, note that $f'(x) = 3x^2 - 2$ has two roots, namely $x \equiv \pm 2 \pmod{5}$, and that

$$f(2) \equiv 4 \pmod{5}, \quad f(-2) \equiv 1 \pmod{5}.$$

Example 136. The elliptic curve $E : y^2 = x^3 + 17$ has bad reduction at $p = 17$, because the $f'(x) = 3x^2$ has the root $x \equiv 0 \pmod{17}$, and

$$f(0) = 17 \equiv 0 \pmod{17}.$$

Example 137. The elliptic curve $E : y^2 = x^3 + 1$ has bad reduction at $p = 3$, because $f'(x) = 3x^2$ and

$$f(2) = 9 \equiv 0 \pmod{3} \quad \text{and} \quad f'(2) = 12 \equiv 0 \pmod{3}$$

(note that in this example any x in $\mathbb{Z}/3\mathbb{Z}$ is a root of $f'(x)$).

Example 138. The elliptic curve $E : y^2 = x^3 - x$ has good reduction at $p = 3$; indeed, $f'(x) = 3x^2 - 1$ has no roots modulo 3, since it is always congruent to -1 modulo 3.

Definition 139. Suppose $E : y^2 = x^3 + ax^2 + bx + c$ is an elliptic curve over \mathbb{Q} , p is an odd prime, and E has good reduction at p . Define

$$E(\mathbb{Z}/p\mathbb{Z}) = \{(x, y) \mid x, y \in \mathbb{Z}/p\mathbb{Z}, \quad y^2 = x^3 + ax^2 + bx + c\} \cup \{\infty\}.$$

Note that $E(\mathbb{Z}/p\mathbb{Z})$ is a finite set, since $\mathbb{Z}/p\mathbb{Z}$ is finite.

Example 140. If $E : y^2 = x^3 + 17$,

$$E(\mathbb{Z}/5\mathbb{Z}) = \{\infty, (2, 0), (3, 2), (3, 3), (4, 1), (4, 4)\}.$$

$E(\mathbb{Z}/p\mathbb{Z})$ forms a group in exactly the same way that $E(\mathbb{Q})$ does. For example, let us find $(2, 0) + (4, 1)$ in the previous example. The slope of the line through these two points is $\frac{1}{2} \equiv 3 \pmod{5}$, so the equation of this line is $y = 3(x - 2) = 3x - 6$. Substituting in the equation for E we obtain

$$(3x - 6)^2 = x^3 + 17$$

$$0 = x^3 - 9x^2 + 36x - 19 \equiv x^3 + x^2 + x + 1$$

$$\equiv (x - 2)(x - 4)(x - 3) \pmod{5}.$$

We find that the third point of intersection with E is $(3, 3)$, and therefore $(2, 0) + (4, 1) = (3, 2)$.

Now let us compute $2(3, 3)$. Implicit differentiation yields $y' = 3x^2/2y$, and

$$y'(3, 3) = 27/6 \equiv 2/1 \equiv 2 \pmod{5}.$$

So the tangent line has equation $y - 3 = 2(x - 3)$, which modulo 5 becomes $y = 2x + 2$. Substituting, we find

$$(2x + 2)^2 = x^3 + 17$$

$$0 = x^3 - 4x^2 - 8x + 13 \equiv x^3 + x^2 + 2x + 3$$

$$\equiv (x - 3)(x - 3)(x - 3) \pmod{5}.$$

We find that our third point of intersection is $(3, 3)$, and hence $2(3, 3) = (3, 2)$. Noting that modulo 5, $(3, 2) = (3, -3)$, we conclude that in $E(\mathbb{Z}/5\mathbb{Z})$, $3(3, 3) = 2(3, 3) + (3, 3) = -(3, 3) + (3, 3) = \infty$.

The previous paragraph illustrates a fundamental difference between $E(\mathbb{Q})$ and $E(\mathbb{Z}/p\mathbb{Z})$. Recall that for $E : y^2 = x^3 + 17$, $(-2, 3)$ is a point of “infinite order” in $E(\mathbb{Q})$, in the sense that the multiples of $(-2, 3)$ form an infinite set. Modulo 5, however, $(-2, 3)$ is the same as $(3, 3)$, and we just saw that $3(3, 3) = \infty$ in $E(\mathbb{Z}/5\mathbb{Z})$. Thus, we might say that modulo 5, $(-2, 3)$ becomes a point of finite order (namely, order 3).

Recall that $2(3, 3) = (3, 2)$ in $E(\mathbb{Z}/5\mathbb{Z})$ and $2(-2, 3) = (8, -23)$ in $E(\mathbb{Q})$; modulo 5, $(3, 2)$ and $(8, -23)$ are the same. Furthermore, we saw that $3(3, 3) = \infty$ in $E(\mathbb{Z}/5\mathbb{Z})$; note that $3(-2, 3) = (19/25, 522/125)$ in $E(\mathbb{Q})$. In general it is the case that if p is a prime of good reduction for E , $P \in E(\mathbb{Q})$ and p is a divisor of the denominators of the coordinates of nP , then $nP = \infty$ in $E(\mathbb{Z}/p\mathbb{Z})$. For example, the denominators of the x - and y -coordinates of $4(-2, 3)$ in $E(\mathbb{Q})$ are 23^2 and 23^3 , and therefore $4(-2, 3) = \infty$ in $E(\mathbb{Z}/23\mathbb{Z})$. Note also that since $3n(3, 3) = \infty$ in $E(\mathbb{Z}/5\mathbb{Z})$ for every $n \geq 1$, the denominators of the coordinates of $3n(-2, 3)$ in $E(\mathbb{Q})$ will be divisible by 5 for every $n \geq 1$.

Definition 141. If E is an elliptic curve defined over \mathbb{Q} and p is a prime of good reduction for E , let $a_E(p) = p + 1 - \#E(\mathbb{Z}/p\mathbb{Z})$.

Example 142. Recall that for $E : y^2 = x^3 + 17$,

$$E(\mathbb{Z}/5\mathbb{Z}) = \{\infty, (2, 0), (3, 2), (3, 3), (4, 1), (4, 4)\}.$$

Hence $a_E(5) = 5 + 1 - 6 = 0$.

The reason for defining the quantity $a_E(p)$ is as follows. If $r, s \in \mathbb{Z}/p\mathbb{Z}$, then $(r, s) \in E(\mathbb{Z}/p\mathbb{Z})$ if and only if $s^2 \equiv f(r) \pmod{p}$. Thus r is the x -coordinate of an element of $E(\mathbb{Z}/p\mathbb{Z})$ if and only if $f(r) \equiv 0 \pmod{p}$ or $f(r)$ is a quadratic residue modulo p . Now, f has at most 3 roots by Theorem 72, so let us focus on the latter case. If $f(r)$ is a quadratic residue and $s^2 \equiv f(r) \pmod{p}$, then by Proposition 74 there are exactly two elements of $E(\mathbb{Z}/p\mathbb{Z})$ having x -coordinate r , namely $(r, \pm s)$. We know by Proposition 88 that exactly half of $\{0, 1, 2, \dots, p-1\}$ are quadratic residues, and it is therefore natural to expect that $f(r)$ will be a quadratic residue for about half of the p possible values of r . Since we get two corresponding y -coordinates for each of the values of r producing a quadratic residue, we would expect to produce roughly $2(p/2) = p$ elements of $E(\mathbb{Z}/p\mathbb{Z})$ in this way. Throwing in ∞ , then, we expect that $\#E(\mathbb{Z}/p\mathbb{Z}) \approx p + 1$. Thus $a_E(p)$ measures how far off $\#E(\mathbb{Z}/p\mathbb{Z})$ is from this “expected value”.

Notice that if $E : y^2 = x^3 + ax^2 + bx + c = f(x)$ is an elliptic curve over \mathbb{Q} and p is a prime of good reduction for E , then

$$\#E(\mathbb{Z}/p\mathbb{Z}) = \sum_{x=0}^{p-1} \left[\left(\frac{f(x)}{p} \right) + 1 \right],$$

and therefore

$$a_E(p) = - \sum_{x=0}^{p-1} \left(\frac{f(x)}{p} \right).$$